Language Documentation & Conservation Special Publication No. 2

Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas

edited by
Andrea L Berez Jean Mulder Daisy Rosenblum

Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas

edited by

Andrea L. Berez, Jean Mulder, and Daisy Rosenblum


Language Documentation & Conservation Special Publication No. 2

Published as a Special Publication of Language Documentation & Conservation

Language Documentation & Conservation
Department of Linguistics, UHM
Moore Hall 569
1890 East-West Road
Honolulu, Hawai'i 96822
USA

http://nflrc.hawaii.edu/ldc

University of Hawai'i Press
2840 Kolowalu Street
Honolulu, Hawai'i
96822-1888
USA

Cover design by Cameron Chrichton
Cover photograph of salmon drying racks near Lime Village, Alaska, by Andrea L. Berez

Library of Congress Cataloging in Publication data
ISBN 978-0-8248-3530-9

http://hdl.handle.net/10125/4463

Contents

Foreword

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. iii-iv
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4462

This timely special publication represents a welcome stage in the development of the field of linguistics. The discipline is witnessing an intertwining of approaches to the study of language that have been evolving, often as independent strands, for well over a century. Major progress has been made in all traditional domains: phonetics, phonology, morphology, syntax, semantics, and discourse. The importance of understanding how linguistic categories and structures emerge over time, how children acquire them, and how social and cultural uses shape them are now widely recognized. More is being learned every day about how language is utilized by speakers for social and cultural purposes. Technological and analytic advances have opened up possibilities for discoveries unimaginable at the outset, from fine acoustic analysis to the manipulation of corpora. Theories are becoming ever stronger, more detailed, and more sophisticated as we learn more about what occurs in typologically diverse languages. This growth of knowledge has necessarily led to greater specialization on the part of individual researchers: there is simply too much for one person to master. At the same time, it is becoming increasingly clear that no aspect of language can be understood fully in isolation.

There has been a palpable shift in attitudes within the discipline, with greater respect among individual scholars for lines of research outside of their own, and greater appreciation of the contributions each can make toward understanding others. Where once distinctions were sometimes drawn between 'description' and 'theory', there is now general recognition of the fact that the two are necessarily interlocking elements of an empirically-based science. This recognition may have been stimulated, at least in part, by a heightened awareness of the immanent loss of linguistic diversity in the world. Informed documentation and description of endangered languages has become a priority for the field, as one by one, languages are slipping away before our eyes and ears.

The present collection is a fine example of the evolution of the discipline, in particular the fading of impenetrable boundaries. It is now understood that good language documentation and description require a strong theoretical foundation, an understanding of what is currently known about language and at least some of the puzzles that remain. All of the authors represented in this special publication bring such a background to their fieldwork, and each of the articles makes an incisive contribution to linguistic theory. At the same time, it is now generally accepted that any theory of value must be grounded in data. The papers collected here are exemplary in this regard. This volume represents the full range of structural domains, but each study also goes beyond the level of the specific patterns, markers, or constructions under consideration to examine the parts they play in larger domains and their import for larger issues. There is work on the phonetic values of reconstructed vowels as revealed through modern fieldwork; on the tonal foot and the role of low-level allophonic processes in the construction of phonological categories; on prosodic, morphological, and syntactic properties of a range of grammatical and lexical clitics and their implications for a general theory of clitics; on the reconstruction of a complex system of number inflection based on noun class; on non-finite inflection, its interaction with argument structure, and its syntactic function; on the semantic thread of event construal linking seemingly disparate uses of the middle voice; on the spatial, temporal, and discourse functions of determiners; and on the structural, pragmatic, and social functions of discourse markers.

Another fading boundary is that between academic linguists on one side and members of language communities on the other. Work on individual languages, particularly those spoken by smaller populations and those in danger of disappearance, must be shaped not just by current linguistic theory but equally if not more by the needs and desires of communities. Our understanding of linguistic structure would be seriously impoverished without the dedication and insight of the speakers who collaborate in this work. More and more community members are becoming scholars. A recurring theme running through this volume is the important insights provided by speakers.

The papers collected here also reflect a maturing of methodology in the field. There are rich discussions of the interplay between data and theory; between elicitation and unplanned spontaneous speech; between direct observation of speech and the insights of speakers; between qualitative and quantitative findings; between current data and analyses and earlier documentation and description. These threads have always been present in the discipline, but not necessarily within the work of individual scholars. It has long been recognized that linguistics includes elements of hard science, of social science, and of the humanities. It has an empirical basis, with strict standards of responsibility to the data. It is a social phenomenon. It is also a human phenomenon, directly concerned with a product of the human mind. The optimal methodologies for the study of language are not necessarily precisely the same as those for physics, mathematics, psychology, philosophy, or literature. In a sense we can see linguistics coming into its own as a discipline, with its own blend of methodologies. The present special publication provides a fine example of this blend.

Marianne Mithun

Contributors

MELISSA AXELROD, who also specializes in Native American languages and language revitalization, is a professor in the Linguistics Department at the University of New Mexico. Together with Jule Gómez de García and María Luz García, she has been working with Ixil Maya speakers of the Grupo de Mujeres por la Paz of Nebaj, El Quiche, Guatemala, since 2002 on NSF-funded projects to produce an electronic archive of linguistic, cultural and historical materials and, most recently, a grammar of Ixil. Axelrod and Gómez de García have also collaborated on a language revitalization project with the Jicarilla Apache Nation in New Mexico and served as editors on the Dictionary of Jicarilla Apache (2007, Uiversity of New Mexico Press).

ANDREA L. BEREZ is a doctoral candidate in the linguistics department at the University of California, Santa Barbara. She is a descriptive and documentary linguist who works primarily with speakers of Ahtna and Dena'ina, two endangered Athabascan languages of southcentral Alaska. Her linguistic interests include intonation, spatial cognition and discourse-functional approaches to grammar. She is also interested in the development of the technological infrastructure to support language documentation and archiving.

LYNDA DE JONG BOUDREAULT graduated from the University of Texas at Austin where she specialized in descriptive linguistics with particular interest in languages of Mesoamerica. Her research experience includes work on Iquito, a Zaparoan language spoken by about 26 people in the northern Peruvian Amazon, and Sierra Popoluca (also known as Soteapanec), spoken in the southern part of the State of Veracruz, Mexico. In addition to documentary work, Dr. Boudreault is interested in pedagogy and issues of language revitalization and maintenance.

ISRAEL MARTÍNEZ CORRIPIO is a graduate student at the Universidad Nacional Autónoma de México. His main interests include voice systems and argument structure in Mayan languages.

DONNA FENTON studied linguistics at San Francisco State University and UC Berkeley. Her main research interests are language documentation and revitalization, and Mesoamerican linguistics.

MARÍA LUZ GARCÍA is a doctoral candidate in linguistic anthropology at The University of Texas at Austin. Her focus of study is on the use of the Ixil language in creating public collectives. She has been working with Ixil organizations such as the Grupos de Mujeres y Hombres por la Paz, the Comunidades de Población en Resistencia, and community leaders involved in exhumation and reburial of victims of the genocide of Mayas of the 1980s. She has collaborated with Gómez de García and Axelrod on documentation of the Ixil language since 2002.

SPIKE GILDEA is an associate professor of linguistics at the University of Oregon. His research interests include fieldwork and descriptive linguistics of South America, historical linguistics of the Cariban family, grammaticalization, and typological/functional linguistics. He has collected primary linguistic data from speakers of thirteen Cariban

languages: Akawaio (Kapóng), Akuriyó, Apalaí, Arekuna (Pemón), Ikpéng, Hixkaryana, Katxúyana, Makushi, Panare, Patamuna (Kapóng), Tiriyó, Waiwai, Wayana, Xikuyana (Katxúyana), and Yukpa.

JULE GÓMEZ DE GARCÍA is a professor of linguistics in the Liberal Studies Department at California State University San Marcos and a specialist in Native American languages. She has worked on documentation of the Kickapoo and Jicarilla Apache languages and is a co-editor on a dictionary of Jicarilla Apache published by University of New Mexico Press in 2007. With María Luz García, Melissa Axelrod, and the Grupo de Mujeres por la Paz, she has worked on documentation of the Ixil Maya language since 2002. The team is currently collaborating on the production of a grammar of the Nebaj dialect of Ixil, a project funded by the NSF.

BEREND HOFF is a visiting lecturer at Leiden University. He has been engaged in describing the grammar of the Carib language (Kari'nja) since the 1950s, and he also studies language contact, especially the case of Island Carib.

ALESSANDRO JAKER is a graduate student in linguistics at Stanford University. His primary interests are phonetics, phonology, and morphological typology. Recently, his empirical work has been focused on Athabaskan languages, in particular the dialects of Dogrib and Chipewyan spoken in and around Yellowknife, Northwest Territories. His dissertation work centers on the prosodic phonology of Dogrib, and the interaction between verb morphology and the prosody of Dogrib verbs.

OLGA CHARLOTTE LOVICK studies Northern Athabascan languages, in particular Dena'ina and Upper Tanana, both spoken in Alaska. Her main research interest lies in the study of discourse, in particular the study of discourse markers and other particles, and the study of prosody in narrative. Current projects include a bilingual collection of narratives in the Tetlin dialect of Upper Tanana as well as a first foray into the prosody of conversation. Olga received both her Magister and her Ph.D. in General Linguistics from the Universität zu Köln, Germany. Prior to her current position as Assistant Professor at the First Nations University of Canada, she was a Postdoctoral Fellow at the Alaska Native Language Center, University of Alaska Fairbanks.

RICARDO MALDONADO is a professor at The Universidad Nacional Autónoma de México and he is a guest professor at the Universidad Autónoma de Querétaro, Mexico. His research interests center on the relationship between language and cognition with special focus on voice systems, possession and datives of Spanish and Mexican indigenous languages.

SÉRGIO MEIRA is a lecturer in American Indian linguistics as Leiden University's Centre for Linguistics. His research focuses on historical linguistics, fieldwork and description of the Cariban and Tupían language families, as well as language and cognition. He has collected primary linguistic data from speakers of 14 Cariban languages: Akawaio,

Akuriyó, Apalaí, Bakairi, Hixkaryana, Karihona, Kari'nja, Katxúyana, Kuhikuru, Makushi, Tiriyó, Waiwai, Wayana, and Yukpa.

JEAN MULDER is a senior lecturer in the School of Languages and Linguistics at the University of Melbourne. Her research interests range over language documentation, grammatical and discourse analysis, minority language education, and educational linguistics with fieldwork on a variety of languages. She began her study of Sm'algyax by working for five years full time in the field developing the Sm'algyax Language Studies Program in School District No. 52 (British Columbia). She completed her PhD in Linguistics at the University of California, Los Angeles and, prior to her current position, she taught at the University of Victoria, Simon Fraser University, the University of Alberta, and the University of Swaziland in southern Africa.

DAISY ROSENBLUM is a doctoral student in linguistics at the University of California, Santa Barbara. Her work focuses on the documentation and description of American indigenous languages in the Mayan and Wakashan families, with an emphasis on the study of multi-modal interaction. Her research interests include prosody, argument structure, deixis, and grammars of space and time, and the interactions among them. She is currently working with members of the Kwakwaka'wakw Nation in British Columbia to record, transcribe and analyze multiple genres of spontaneous speech in three dialects of Kwak'wala.

HOLLY SELLERS completed a BA (Honors) in Linguistics at the University of Melbourne. She has an interest in syntax as well as undocumented languages. She has worked on data from the Native American language of Sm'algyax from the British Columbia region in Canada as well as first hand documentation of the Ganalbiŋu language of Northern Australia.

LOGAN SUTTON is a Ph.D. candidate in linguistics at the University of New Mexico and works for the American Indian Studies Research Institute at Indiana University. He received his B.A. in linguistics from Indiana University in 2004 and his M.A. in linguistics from the same school in 2006. His primary research interests include synchronic and diachronic analysis and description of the Kiowa-Tanoan and Caddoan language families and language revitalization among Native American communities in the Southwest. He is currently working alongside Professor Melissa Axelrod and other graduate students at UNM in collaboration with members of Pueblo communities towards developing material for language revitalization and maintenance.

Acknowledgments

The editors are deeply grateful to the following individuals who have contributed to the success of this publication: Marianne Mithun, Anthony Aristar, Helen Aristar-Dry, Paul Newman, Roxana Ma Newman, and Martha Ratliff (for their procedural guidance); Stefan Th. Gries (for encouraging the project in the first place); Michal Brody, Matthew Gordon, Salomé Gutiérrez Morales, Gary Holton, Cara Penry Williams, and Sandra Thompson (for sharing their expertise on the contents); Cameron Crichton and Holly Sellers (for their design skills); and Lea Harper, Timothy J.P. Henry, Carmen Jany, Joye Kiester, Verónica Muñoz Ledo, Carlos Nash, Elizabeth Shipley, Rebekka Siemens, and Alex Walker (for being all-around good colleagues).

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 1-8
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4448

1

Introduction: The Boasian tradition and contemporary practice in linguistic fieldwork in the Americas

Daisy Rosenblum and Andrea L. Berez

University of California, Santa Barbara

The papers collected in this volume represent a cross-section of current linguistic scholarship in which authors share a common methodology: the analyses presented here have emerged directly from the challenges and felicities of linguistic fieldwork and language documentation. We take fieldwork to be the study of a language—often not the researchers' own, and frequently conducted on site where the language is spoken—as a holistic system, operating within interdependent social, cultural, and historical contexts. While each contributor in these pages has narrowed his or her focus to one small corner of that system, the work here is informed by attention to the larger contexts of these languages, situated in the communities in which they are spoken, and of field linguistics as a discipline, situated in the community in which it is practiced.

The languages considered here are all indigenous to the Americas. Measured in degrees of latitude, they span nearly half the globe, from Alaska and the Northwest Territories in the north to Brazil in the south. The tradition of fieldwork in American linguistics in particular and in American anthropology more generally was established in large part by Franz Boas in the early twentieth century. Boas and his students, with Edward Sapir foremost among them, located fieldwork as the fundamental method for elucidating universals and variation in language. For even the earliest fieldworkers, gathering data in the field required collaboration with native speakers, the collection of diverse genres of data using a multiplicity of methods, and patience with the unfolding and sometimes chaotic nature of data gathered in the field. In the century since Boas conducted his iconic field trips to North America's Northwest Coast, generations of Americanist linguists have carried forward this legacy, and the fifteen authors in the present publication carry forth this tradition. The articles coalesce around these same themes of community collaboration, attendance to a wide range of data types and methods, and the slowly unfolding nature of the process. Each paper highlights the methodological and theoretical contributions field-based linguistic research has made to our understanding of grammatical structure in American languages, and to the broader theoretical objectives of the discipline as a whole. To paraphrase Jean Mulder and Holly Sellers (chapter 3), contemporary linguistic theories can give insight into the underlying structure of field-collected language data in typological context, and the data themselves afford empirical means of testing and expanding our theories.

COMMUNITY COLLABORATION

> A general review of our ethnographic literature shows clearly how much better is the information obtained by observers who have command of the language, and who are on terms of intimate friendship with the natives, than that obtained through the medium of interpreters. (Boas 1911 [1966, 1991]:57)

A common starting point for a field linguist is a mutually beneficial relationship with the community of speakers with whom one is to work; whether a linguist is a member of the speaker community or a guest, a reciprocal and collaborative alliance is crucial to success. An exchange of skills and expertise between speakers and linguists has been integral to successful long-term fieldwork relationships since Boas collaborated with George Hunt to transcribe Kwak'wala (Kwakiutl) texts and produce ethnographies of Kwakwaka'wakw communities. Many fieldworkers see this type of exchange as a way of giving back to speakers who are sharing their linguistic and cultural knowledge.

However, a collaborative relationship between speakers and linguists is more than merely responsible procedure; it produces better data and can illuminate patterns otherwise inaccessible to researchers. In her 2001 essay reflecting on the roles of speakers and linguists in language documentation, Mithun observes that

> [i]n many ways, the more the speaker is invited to shape the record, the richer the documentation of the language, and the more we will learn about the extent to which languages can vary … If speakers are allowed to speak for themselves, creating a record of spontaneous speech in natural communicative settings, we have a better chance of providing the kind of record that will be useful to future generations. (Mithun 2001:51-3)

Several of the papers in this publication address how such relationships are established and negotiated, as well as the benefits to research, analysis, and theory deriving from close collaboration with speakers.

Jule Gómez de García, Melissa Axelrod, and Maria Luz García (chapter 2) present their analysis of the development and use of an Ixil Maya discourse particle against the backdrop of six years of collaboration with a women's weaving and agricultural cooperative in Guatemala. It is the dynamic of the group itself and the women's metatextual conversations about language work that generated the analysis presented here. In the midst of transcription session discussions, the event-sequencing function of the particle came to light as members of the group, unaccustomed to the picture book stimulus, realize the sequentiality of *Pancakes for Breakfast* and *Frog, Where Are You?*. "One of the most relevant facts of our fieldwork in Nebaj is the varying levels of literacy skills among the women and the correspondence of those skills with the differentials in experience the women have had with printed materials. This fact has determined how transcription and elicitation sessions will be arranged and now it is clear that it is also a factor in shaping how, and how frequently, the discourse marker *vet* is used by the women in their speech" (p. 29).

In turn, Jean Mulder and Holly Sellers's paper on classifying the uncommonly wide range of clitics found in Sm'algyax (Coast Tsimshian; chapter 3) shows how Anderson's 2005 constraint-based account of clitics provides a framework for understanding the com-

plex behavior of Sm'algyax clitics. In addition, the Sm'algyax clitics provide insight into the applicability of the theory. Many of the examples cited come from texts that have been painstakingly edited by a collective of Sm'algyax speakers and writers. "This is a case where not only does linguistic theory help sharpen our understanding of fieldwork data, but also where field linguistics has consequences for linguistic theory," they write, "… [t]he motivation for approaching fieldwork and theoretical linguistic analysis in this way is to work toward enabling knowledge of the language to be constructed not only for and with, but also by community-members" (p. 34-35).

Logan Sutton (chapter 4) tackles the role of collaborative fieldwork in historical and comparative linguistics in his reconstruction of noun class and number in Kiowa-Tanoan languages. At the same time he acknowledges the duty field linguists have toward respecting a language community's desire for privacy. Pueblo communities have a long history of protecting certain aspects of culture and tradition, including language, from academic research by outsiders. Sutton, whose research involves collaboration with several Native American communities of the southwest United States, justifiably limits his historical analysis here to data already in print. He writes that in a publication about the relationship between fieldwork and analysis, it may "be surprising that I base the hypothesis and the conclusions of this paper on data collected by somebody else. … However, with all the virtues of fieldwork—indeed, most linguistic work is ultimately owing to native speakers sharing their languages—there are restrictions that must be recognized and respected" (p. 58). The sensitive and often emotionally-charged issue of privacy is a daily reality for many language workers in the Americas and elsewhere.

A DIVERSITY OF RESOURCES

> To plumb the depths of a language, all sources are of value—elicitation, texts, casual speech, stories, and conversations … diversity must be allowed to suffuse fieldwork. (Rice 2001:240-1)

When linguists go to the field to gather data, the kinds of data they collect, the manner in which they are collected, and the way in which these data are processed are primary considerations with far-reaching theoretical implications. Fieldworkers must decide when to collect data through elicitation or recording of spontaneous discourse; what paradigms, texts, and lexica they seek; which orthography or orthographies they will choose for transcriptions, and whether recordings are edited or left in their original state. A diversified approach to linguistic data, and faith in the long-term theoretical promise of such data, is also part of the Americanist tradition. Mithun points out that in addition to direct elicitation, "[a] second kind of methodology, the recording of connected speech, formed the core of much linguistic fieldwork over the past century, particularly in North America. The tradition of text collection arose in part from a desire to document the rich cultures of the speakers, but it was also seen as a tool for understanding languages in their own terms, rather than through European models. The texts served as the basis for grammatical description" (Mithun 2001:35). She quotes Boas' introduction to the inaugural volume of the *International Journal of American Linguistics*, in which he presents a mandate for a new era in the study of American languages and language in general: "While until about 1880 investigators

confined themselves to the collection of vocabularies and brief grammatical notes, it has become more and more evident that large masses of texts are needed in order to elucidate the structure of the languages" (Boas 1917:1, quoted in Mithun 2001:35).

Several papers here address the value of collecting and drawing on a diversity of resources, from legacy materials in (and out of) archives to newly recorded speech, and the importance of bringing fresh interpretation to the data and theories we may inherit from others. The chapter by Spike Gildea, B.J. Hoff, and Sérgio Meira (chapter 5) is also a historical reconstruction, this time of the proto-vowel *ô* in Cariban languages. Here, however, the focus is on how the modern collection of reliable field data can help answer questions remaining in the fieldnotes of our academic predecessors. Special attention is paid to how modern linguists can benefit from the intervening years of professional and methodological growth the field has seen since the early days of documentary work. "Early sources of data from Cariban languages were vexed with poor transcription, especially of vowels. As a result, early attempts to compare Cariban wordlists resulted in inconsistent correspondences. In our own fieldwork, we have found that old word lists are rarely confirmed by our modern transcriptions" (p. 93). "This paper illustrates the importance of good modern fieldwork in two ways: first, after collecting reliable modern data, many inconsistencies in cognate sets disappear and previously unseen patterns become clear; second, reliable modern data can provide insight into previously opaque transcription systems used by older sources, thereby enabling us to make better use of language data recorded several hundred years ago" (p. 92).

The importance of looking to a combination of connected spontaneous speech and elicited data is addressed by Fenton, Lovick, and others, in particular the way in which multiple types of data examined together can illuminate the rich complexity of a linguistic system. Olga Charlotte Lovick builds on the vast body of work on clause-level syntax in Athabaskan languages in her analysis of several discourse markers in Dena'ina (chapter 8). She invokes a range of data sources—e.g., fieldnotes and texts she has collected and those collected by her predecessors—and methods to uncover the functions of particles that, according to speakers, "'have no meaning' or 'mean something else in every sentence'" (p. 174). The difficulty of pinning down the functions of discourse particles before lexical and syntactic analyses are available is clear from the fieldnotes of previous researchers Lovick consulted, which are teeming with particles that are either variably glossed or not glossed at all. In her chapter, Lovick advocates a flexible approach for revealing the connective and cohesive functions of each of the particles she examines, and allows both direct elicitation and analysis of narrative to be her guide. "To define the meaning of one of these discourse markers, simple elicitation is not sufficient … Direct questioning about the meaning of the particles considered here may or may not yield a useful answer" (p. 174). She continues, "only discourse analysis can show that [the particles] have different functions corresponding to their position within a narrative unit" (p. 174).

In their chapter on middles and reflexives in Yucatec Maya (chapter 7), Israel Martínez Corripio and Ricardo Maldonado also advocate a multi-pronged approach to fieldwork-based linguistic analysis, including elicitation, text analysis, and, most importantly here, heeding speaker intuition as a clue to the semantic features of the constructions under investigation. Despite the seemingly arbitrary distribution of middles and reflexives in Yucatec Maya, the authors show that middles are limited to *absolute* events (i.e., those in

which no energy is expended), a typologically unusual motivation among languages with middle systems. An essential component of their methodology was the incorporation of speaker intuition in the analytic process: "[o]ur data collection began with direct elicitation and the analysis of oral narrative, but these—whether alone or considered together—were not sufficient to fully illuminate the behavior of the YM middle system. As our analysis grew, we found it necessary to invent ways to investigate speaker intuition as well" (p. 148). By trusting speakers' intuitions about subtle semantic differences between hypothetical language-use situations, Martínez Corripio and Maldonado are able to illuminate and confirm a coherent semantic domain governing the use of middles and reflexives in Yucatec.

Another paper highlighting the role of multiple methodologies in fieldwork is Donna Fenton's study of determiners in Teotitlán del Valle Zapotec (chapter 6). The determiners have a number of temporal, discourse and spatial functions, the details of which only became apparent through text analysis and elicitation, each method revealing a different set of subtleties of use. "The texts, which owing to the circumstances of my work in Teotitlán del Valle were collected and transcribed in the early stages of the research, revealed temporal and discourse functions of the determiners. Elicitation, which was informed by the analysis of the texts, brought out the expected spatial distinctions and enabled me to refine my analysis of the temporal extensions of the determiners. Neither method on its own would have provided a complete picture, and it is possible that the 'backwards' order of data collection allowed for a better insight into how speakers actually used and thought about the system of determiners" (p. 126).

Importantly, as noted by Mithun, a multiplicity of approaches will "permit … us to notice distinctions and patterns that we might not know enough to elicit … this material is in many ways the most important and exciting of all. Linguistic theory will never be moved ahead as far by answers to questions we already know enough to ask as it will by discoveries of the unexpected" (Mithun 2001:45). As our work as linguists now finds application among communities eager to maintain and revitalize their languages, the need for a diversity of data is especially pressing.

EVOLUTION OF THE ANALYSIS

> Fort Rupert, September 22, 1886
>
> The material overwhelms me. I am very curious about the story I am to get today. Each day brings so much that is new and so many surprises. (Boas, in a letter to his parents, from Rohner 1969:24)

> Ladners Landing, August 17, 1890
>
> With what I am getting now and what I got before, I can see that I have learned a great deal. Much of the confused knowledge I had then is being straightened out now. (Boas, in a letter to his wife, from Rohner 1969:65)

Often, what turns out to be the most valuable data may only reveal itself as such after the fieldtrip is over and the researcher has returned home. This revelation may eventually lead to another trip to the field—and another. In all cases, though, data gathered through fieldwork is chaotic, messy and full of surprises. Understanding it may happen only slowly or in fits and starts, often over multiple field seasons. This facet of linguistic fieldwork requires patience, curiosity, flexibility, and a focus on the process of discovery.

In many of the essays gathered here, like Fenton's analysis of Zapotec determiners discussed above, the analysis evolved over time through multiple visits to the field site. This theme is also present in Lynda Boudreault's description of dependent verbs in Sierra Popoluca (chapter 10). The author describes dependent verbs in the language and examines the particular constructions in which they appear. The analysis presented here began with a careful look at previous work and developed by means of recurrent text transcription and elicitation that took place over multiple fieldtrips between 2004 and 2007. It is the inherently cyclical nature of field-based linguistic analysis that Boudreault highlights: "text transcription, data mining, post-hoc analysis, and controlled elicitation … is a cyclic process that looks to data to corroborate predictions driven by linguistic theory and that in turn bear on theory" (p. 256). Her goal here is to show "the interdependent processes of data collection and analysis by addressing how different observations emerge at different stages of the analysis" (p. 226).

Alessandro Jaker, in his paper on gemination and tonal feet in Weledeh Dogrib (chapter 9), also focuses on linguistic analysis as a process that develops over time, but in this case the evolution is across generations of linguists. Each new linguist to tackle an issue inherits the assumptions and traditions of his or her academic predecessors, but at the same time brings something uniquely personal to the analysis. It is up to the resourceful linguist to draw upon his or her own experiences to move the field forward. As Jaker notes, American structuralists have inherited from Bloomfield the notion that consonant length is a phonemic issue and therefore must be a feature that can distinguish between utterances. Consonant length in Athabaskan languages is often seen as phonetic, not phonemic (see Jaker for references). In Dogrib, however, the picture of consonant length is more complex. Jaker shows that in light of the Optimality Theory analysis presented here, "the key generalizations about morphophonemics in Dogrib require reference to syllable weight, which in turn requires reference to consonant length—we would miss important generalizations if geminates were not included as part of the phonology" (p. 204). Jaker brings his intuitions as a native speaker of another language with contrastive consonant length, Italian, to his fieldwork, which he calls "a process of unraveling layers of unstated assumptions, both others' and my own. Relying on descriptive statements made by others means adopting their assumptions about what facts ought to be included in the description and what should be thrown out. Fieldwork with speakers of Weledeh Dogrib enables me to go back to the original speech signal and decide for myself what is structurally important and what is not" (p. 204).

This collection takes its place alongside several recent publications in which field linguists share expertise and anecdotes about the particular methodologies that make fieldwork, and the scholarly analyses based on field-collected data, so compelling (Bouquiaux & Thomas 1992; Vaux and Cooper 1999; Newman & Ratliff 2001; Gippert et al. 2006; Crowley 2007;

Bowern 2008; Thieberger forthcoming). We offer the current discussion to readers with interests in the languages of the Americas and beyond. The themes considered in these papers are in many ways pertinent for field linguists worldwide.

References

Boas, Franz. 1911. Introduction to *Handbook of American Indian Languages* (Bulletin 40, Part 1, Bureau of American Ethnology), 1-79. Washington: Government Printing Office, 1-83. Reprinted 1966 [1991]. Lincoln: University of Nebraska Press.

Boas, Franz. 1917. Introduction. *International Journal of American Linguistics* 1(1).1-8.

Bouquiaux, Luc and Jacqueline M. C. Thomas. 1992. *Studying and describing unwritten languages.* Dallas: Summer Institute of Linguistics.

Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide.* New York: Palgrave Macmillan.

Crowley, Terry. 2007. *Field linguistics: A beginner's guide.* Oxford: Oxford University Press.

Gippert, Jost, Nikolaus P. Himmelmann, and Ulrike Mosel (eds.). 2006. *Essentials of language documentation.* Berlin, New York: Mouton de Gruyter.

Mithun, Marianne. 2001. Who shapes the record: The speaker and the linguist. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 34-54.

Newman, Paul and Martha Ratliff (eds.). 2001. *Linguistic fieldwork*. Cambridge: Cambridge University Press.

Rice, Keren. 2001. Learning as one goes. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 230-249.

Rohner, Ronald P. 1969. *The ethnography of Franz Boaz, letters and diaries of Franz Boas.* Chicago: University of Chicago Press.

Thieberger, Nicholas (ed.). Forthcoming. *The Oxford handbook of linguistic fieldwork.* Oxford: Oxford University Press.

Vaux, Bert and Justin Cooper. 1999. *Introduction to linguistic field methods.* Munich: LINCOM.

Daisy Rosenblum
drosenblum@umail.ucsb.edu

Andrea L. Berez
aberez@umail.ucsb.edu

2

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 9-31
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/101025/4449

Sociopragmatic influences on the development and use of the discourse marker *vet* in Ixil Maya

Jule Gómez de García[a], Melissa Axelrod[b], and María Luz García[c]

[a]*California State University, San Marcos*
[b]*University of New Mexico*
[c]*The University of Texas at Austin*

In this paper we explore the functions of the particle *vet* in Ixil Mayan and argue that it is a discourse marker used to perform both structural and pragmatic functions. *Vet* serves as a structural marker indicating temporally or causally interdependent items; it also has sociopragmatic functions, allowing speakers to present an evaluation of a discourse that invites interlocutors to also take a stance both on the information presented and on their roles in particular sociocultural activities. These functions of managing negotiations among interlocutors range from agreements on descriptive terms to calls for social action among entire groups, in all cases highlighting the social nature both of discourse and of group activity. The overlapping of the structural and pragmatic functions of *vet* demonstrates the grammaticalization cline ranging from adverb to discourse marker proposed by Traugott (1997). Our examination of *vet* in a range of genres produced by the *Mujeres por la Paz* of Nebaj, El Quiché, Guatemala, a cooperative formed in 1997 by Ixil Maya women who were widowed or left fatherless during the Guatemalan civil war, suggest that the effects of the individual and group identities and motivations of participants outweighs anticipated genre effects.

1. INTRODUCTION. The twenty-nine members of the *Grupo de Mujeres por la Paz* live in Nebaj, El Quiché, Guatemala, a bustling town in the Cuchumatanes Mountains, seven hours by bus from Guatemala City. The *Grupo* is an unofficial agricultural and weaving cooperative formed in 1997 by Ixil Maya women who were widowed or left fatherless during the Guatemalan civil war. In the early 1980s, over four hundred highland Mayan villages, many in the Ixil area, were destroyed and over a million people were displaced within the country or forced to move into Mexico. Thousands of them took up hiding in refugee communities in the mountains where they lived until the mid-1990s in a state of constant flight from government troops (Sanford 2003, Falla 1998, Manz 1988). Many women returned to Nebaj from the mountains with their children and formed cooperatives such as the *Grupo de Mujeres por la Paz* to provide themselves with support and sustenance as they began rebuilding their lives and as they struggled to support their families.

The members of the *Grupo de Mujeres por la Paz* have been working with us for six years to produce a multimedia corpus of texts, photographs, videos and audio files that document their wartime experiences and their efforts at recovering or reconstituting their

communities and families in the post-war years of the past decade.[1] Their recorded narratives describe the progress of the *Grupo*'s weaving and agricultural cooperative, and they document the women's emerging literacy. In 2002, they agreed to help with the transcription of the audio portion of a video of their initial planning meeting for the construction of a greenhouse. At that time, we established a process that continues to work well for all of us. As there are three of us on the academic team, we invited the women to come to elicitation and transcription sessions in groups of three, a format with which they have been quite comfortable. Patiently, they participated in the work of *repitiendo palabras*, repeating what they heard on tape slowly so that we could write it down. In order to more fully participate in the transcription work, they decided they wanted to learn to read and write Ixil. Two members of the *Grupo* who are proficient writers of both Spanish and Ixil organized biweekly classes where the others learned to hold and manipulate pencils and pens, then to sign their names, then to write the Ixil alphabet and, eventually, words and phrases. Then they hired a professional teacher and took first-grade general education classes which included not only literacy lessons, but also civics, math, and general health studies.

The women took control of the transcription sessions, learned to use computers to play back audio files, and worked to transcribe their own narratives, to put their own words into writing. Currently (Spring 2008), some of them are learning to use computers to type the transcriptions from the notebooks they have painstakingly produced over the past two years. This entire process was carefully video or audio recorded, as were a great number of meetings, transcription sessions, personal histories, story-book narrations, and, of course, parties.

As we transcribed, bilingual members of the *Grupo* provided translations into Spanish, translations that were often negotiated by the women who began to wonder at the functions of individual words. These discussions led to the general interest in grammar and, specifically, to our investigation of the particle *vet.* The recordings gathered for the multimedia database provide the data for this paper and the many hours the women spent in negotiating meaning provide the impetus and the context for analysis.

2. DISCOURSE MARKERS. As we were observing the members of the *Grupo de Mujeres por la Paz* in a multitude of social contexts where interpretation of the purpose of their activities and the negotiation of process were the conversational tasks, we began to focus on linguistic elements used by the women to accomplish those tasks. In this paper we explore the functions of the particle *vet* and argue that it is a discourse marker used to perform both structural and pragmatic functions.

Fraser (1999:950) defines discourse markers as

> a pragmatic class, lexical expressions drawn from the syntactic classes of conjunctions, adverbials, and prepositional phrases. With certain exceptions, they signal a relationship between the segment they introduce, S2, and the prior seg-

[1] This project is supported by the National Science Foundation, BCS Grant Nos. 0504804 and 0504905. This work would not be possible without the dedication, patience, and good will of the members of the *Grupo de Mujeres por la Paz,* to whom we are extremely grateful.

> ment, S1. They have a core meaning which is procedural, not conceptual, and their more specific interpretation is 'negotiated' by the context, both linguistic and conceptual.

Schiffrin also defined discourse markers in terms of their use to link segments of discourse. She called them "sequentially dependent elements that bracket units of talk" (1987:31) and listed such defining characteristics as syntactic detachability and occurrence in initial position as well as their ability to occur with a range of prosodic contours and 'levels' of function within the discourse. Her later work (Schiffrin 2001:66) highlights the fact that they are multifunctional elements that also index the discourse to the discourse participants. They are

> indices of the underlying cognitive, expressive, textual, and social organization of a discourse … [and that it is] ultimately properties of the discourse itself (that stem, of course, from factors as various as the speaker's goals, the social situation, and so on) that provide the need for (and hence the slots in which) markers appear.

Redeker describes this indexing function as one that serves to bring to the listener's attention "a particular kind of linkage of the upcoming utterance with the immediate discourse context" (Redeker, 1991:1168). Thus, in addition to linking propositions, that is, to providing coherence within a discourse, discourse markers have an important metatextual function of linking the discourse to its producers and interpreters. As Traugott (1997) puts it, discourse markers "allow speakers to display their evaluation not of the content of what is said, but of the way it is put together; in other words, they do metatextual work." Crismore et al (1993) describes metatextual devices as strategies that allow speakers or writers to draw attention to the "act of discoursing," and help interlocutors to organize, classify, interpret, and evaluate propositional material.

It is this metatextual function of discourse markers that allows manipulation of the alignments and of the roles and responsibilities of participants in the interaction. Discourse markers used metatextually allow speakers to present an evaluation of a discourse that invites interlocutors to also take a stance regarding the information presented.

What we will see in the analysis of the data is the duality of function of this discourse marker. It serves as a structural marker indicating temporally or causally interdependent items; it also indicates metatextual comment on the discourse itself and, importantly, on the interlocutors' roles in particular sociocultural activities. We find that the occurrences of *vet* thus include structural adverbial functions that often co-occur with the sociopragmatic functions of managing negotiations among interlocutors. These sociopragmatic uses range from agreements on descriptive terms to calls for social action among entire groups, in all cases highlighting the social nature both of discourse and of group activity. This use of *vet* is similar to the notion of indexical 'shifter', which Hanks (1987:682) defines as items that "relate utterances to their speakers, addressees, actual referents, place and time of occurrence. Indexical centering is a primary part of the interpretation of discourse because it connects the evaluative and semantic code with the concrete circumstances of its use." The overlapping of the structural and pragmatic functions of *vet* demonstrates the grammaticalization cline ranging from adverb to discourse marker proposed by Traugott (1997).

3. ADVERBIAL USES OF VET. The only previous analysis of *vet* is that of Ayres (1991), who focuses on the use of *vet* as an adverbial that "means something like *ya* ['already'], *entonces* ['then'], or *luego* ['soon, later']. It is used after the verb, before the directional that follows the verb" (our translation from the original Spanish).[2] Ayres' definition indicates that adverbial *vet* encodes action concluded prior to the speaking moment, or an action or event that occurs concurrently with or sequential to another event, as in the following examples from Ayres (1991:150):

(1) *Kat taq'cha **vet** ok.*
'**Ya** lo llevó otra vez para adentro.'
'He's already taken it inside again.'

(2) *As max inima **vet** unq'a mama'la tenam u Jesuus ta'n viyol chajnaj.*
'Entonces muchos del pueblo creyeron en Jesús por las palabras de ellos.'
'Then many in the town believed in Jesus because of his words.'

According to England (2007), aspect or mood markers are used in Mam, a closely related Mayan language, to signal a shift from the time frame of the narrative to the actual moment of speech in a commentary on the narrative or to signal the narrative's completion. Time, she says, "is inferred from aspect, unless it is indicated directly by time adverbs, which are, however, not very frequent. … Temporal overlap or simultaneity is shown by the use of conjunctions" (2007:20). Ixil likewise relies heavily on the use of aspect and subordinating conjunctions to indicate temporal sequencing while having relatively few temporal adverbs. Nonetheless, *vet* occurs quite frequently in Ixil discourse, a fact which we attribute to its double role as both an adverb and a discourse marker. We will suggest that this pattern of use of *vet* reflects a process of grammaticalization of *vet* from adverb to discourse marker currently taking place in the language.

4. DATA AND METHODOLOGY. We began our analysis with a simple count of occurrences of *vet* in four texts in our data set representing three genres. Two of the texts were retellings of the captionless stories *Frog, Where are You?* (Mayer, 1969) and *Pancakes for Breakfast* (diPaolo, 1978). The other two texts were transcriptions of the proceedings of two business meetings of the full membership of the *Grupo de Mujeres por la Paz*, the first a speech at a meeting and the second a portion of discussion between meeting participants. The use of data from different genres of discourse was key to our analysis. As Hanks (1987:670) points out, "viewed as constituent elements in a system of signs, speech genres have value loadings, social distributions, and typical performance styles according to which they are shaped in the course of utterance." We anticipated that differences between the styles of discourse we examined would be reflected in differences in the use of *vet*, but as we will show, genre was only one of the factors involved in the distribution of

[2] "Esta palabra significa algo como *ya*, *entonces*, o *luego*. Se usa después del verbo, antes del lugar del direccionl que sigue al verbo" (Ayres 1991:150).

the discourse marker *vet*. Hanks (1987:687) argues that "discourse genres are part of the linguistic *habitus* that native actors bring to speech, but that such genres are also produced in speech under various local circumstances." The data examined here revealed a wealth of such local circumstances that altered our initial expectations about genre in accounting for the use of *vet*.

Definitions of discourse markers commonly refer to their use in linking units or segments of discourse, so it is appropriate here to comment on the nature of these kinds of units. Discourse units have been defined in a variety of ways in the literature. Units of analysis have included grammatical units, such as clauses or phrases, and also intonation units (IUs). IUs do not necessarily coincide with grammatical structures and are usually defined by such prosodic cues as pauses, acceleration and deceleration, pitch levels, terminal pitch contours, and voice quality (Chafe 1994:58-9). According to Chafe (1994:63-4), intonation units can be *fragmentary*, or truncated, or they can be successful; these latter units can be classified as either *substantive*, conveying "ideas of events, states or referents," or *regulatory*, serving to regulate "interaction or information flow." Other candidates for discourse units that appear in the literature include turns (e.g., Ford, Fox, and Thompson 2002), semantic/pragmatic propositions (e.g., Fraser 1999), and narrative episodes (e.g., Johnstone 1990, 2002).

Our analysis is based on oral interactions which have been transcribed in lines corresponding to IUs, and we have examined all material on both sides of each instance of *vet* in the transcriptions. This means that our analysis considers the local context for *vet*, within the IU and across adjacent IUs, as well as the more global context across the wider span of the discourse (Schiffrin 2001:57).

We transcribed, analyzed, and numbered the IUs in the texts. We then made a record of all of the occurrences of *vet* in IUs 101-200 in each text, skipping over the first hundred IUs which contained much introductory or 'set-up' information that might not genuinely reflect the generic usage of *vet*. Counting the occurrences in those hundred IUs of each text yielded the results in Table 1:

Table 1: Occurrences of *vet* in clauses 101-200 in four texts

Frog, Where are You?	26
Pancakes for Breakfast	4
Literacy Meeting	12
Thesis Meeting	5

We determined from this small sample that the frequency of usage of *vet* is not predictable by genre alone and that other factors in the discourse situation influenced the use of this discourse marker.

Given Ayres' analysis of *vet* as an adverb, our expectation was that we would find speakers using *vet* to advance the plot in story retellings where the order of activities is essential to moving a plot forward. For other genres, our hypothesis was that adverbial *vet* would serve a corresponding function of ordering discourse units within a larger discourse.

We also expected that *vet* would take on the functions of indexing the discourse and its participants to understood and expected social organization in meetings and discussions where negotiation required such a linkage. In other words, we expected the use of *vet* to vary according to genre, with stories containing more instances of *vet* with structural and adverbial functions only and other talk containing more instances of *vet* with pragmatic functions. Our hypothesis was that the use of *vet* would be linked to particular syntactic and discourse functions that meet particular pragmatic needs of the speakers in each context.

We begin our discussion with an example of the kind of storytellings that we had expected would contain the most adverbial and structural uses of *vet*: narratives based on the stories depicted in the captionless books *Frog, Where Are You?* (Mayer 1969) and *Pancakes for Breakfast* (diPaolo 1978). We will see that there are indeed adverbial uses of *vet* in the narrative based on *Frog*. The uses of *vet* in *Pancakes* are particularly interesting, as they vary in function according to the women's understanding of the book: the uses of *vet* are more metatextual in nature until the women realize that the pictures represent a narrative structure. At that point, we see an increase in the use of *vet* to signal that structure.

In Sections 4.3 and 4.4, we examine two other genres: a speech by one woman to the group, and a sample of discussion at a meeting.

4.1. Occurrences of *vet* in a *Pancakes for Breakfast* narrative. The retelling of *Pancakes for Breakfast* we analyze here is very typical of many of the Ixil retellings of this book in that there is very little use of the expected use of *vet* to provide coherence or to signal sequencing in the first seventy of the hundred IUs selected for analysis, as reflected in Table 1 above. This analysis comes from a session in which the three speakers, Nan R,[3] Nan Xh, and Nan P, were baffled by the pictures in the book presented to them and struggled to interpret and then describe what they were seeing, clearly unaware that a story could be extracted from these pictures. *Pancakes* is the story of an older woman who decides to make pancakes one morning but realizes that she does not have all of the ingredients she needs. She goes to her barn and fetches some eggs and then, when she finds she has no milk, has to return to the barn to milk her cow. She has no syrup so she leaves her pancake batter and walks to a shop to buy some. While she is gone, her cat and her dog spill the batter on the floor and she returns to find the mess that means she will not have pancakes. But the smell of pancakes wafts to her from her neighbors' house and she goes to have breakfast with them.

The group began their 'narrative' by pointing to and discussing the things they saw in the pictures, naming the dog and the cat, asking about an item they could not identify, a dress hanging from a coat rack on a hanger. They interpret a washbasin full of soapy water as a basket of tortillas, presumably because the thought bubble over the woman's head contains a stack of pancakes that have the size, shape, and color of a stack of fresh corn tortillas. They are not sure at first if this is a man or a woman, but upon seeing her earrings decide that she is a woman, study the picture a bit longer and proclaim that she is pregnant.

---

[3] *Nan* is an Ixil honorific used when addressing or talking about a woman who is middle-aged or older. In our texts, we have assumed this convention of using *Nan* plus the initial of a woman's first name to identify speakers.

One of the researchers suggests that she is an older woman and, looking carefully again, they agree that she is old and therefore not pregnant but fat. One member of the group, Nan P, ends this discussion by first stating in sentence (3a) that the woman is old, then explains immediately in (3b) how she recognizes this. The women continue describing the activities they see in the pictures.

(3) (Nan P)

a. *Qestu chit vete'.*
qestu chit **vet**-e'
old very DM-FV[4]
'Está grande (de edad) ya.'
'She's very old.'

b. *Qestu chit vet aaki ni vile'.*
qestu chit **vet** aaki n-i v-il-e'
old very DM person.of.respect.FV INCOMP-EV 1S.ERG-see-FV
'Yo veo que está anciana.'
'I see she's old.'

Notice the use of *vet* in sentences (3a) and (3b). Sixty-four IUs of the one hundred IUs analyzed intervene between the use of *vet* in (3b) and the next use of *vet* in (4a-c) below. The intervening clauses are further attempts at decoding five pages of pictures. Eventually, one of the women, Nan R, refers to the woman in the pictures as a *patoja* 'young girl' (4a), immediately asks herself if it is a girl or a woman (4b), and finally responds to her own question (4c) that the woman is old. The use of *vet* in (4c) links her comment to the previous discussion of the woman's age.

[4] The first line of each example represents an Intonational Unit of spoken Ixil, the second line contains the morphological parsing, the third contains grammatical category labels for each morpheme, the fourth is a Spanish translation provided by the speakers, and the fifth is an English translation provided by the authors. Abbreviations used: 1P.ABS='1st person plural absolutive'; 1S.ERG='1st singular ergative'; 1P.ERG='1st plural ergative'; 3S.ERG='3rd person singular ergative'; 3P.ERG='3rd person plural ergative'; COMP='completive aspect'; CPZ='complementizer'; DM='discourse marker'; DET='determiner'; DET.PL='plural determiner'; DIR='directional'; EMP='emphatic'; EV='epenthetic vowel'; FOC='focus particle'; FV='final vowel'; INCOMP='incompletive aspect'; INST='instrument'; LOC='locative'; NEG='negative'; PART='particle'; PAS='passive'; PL='plural subject'; POT='potential aspect'; PP='past participle'; REL='relativizer'; RN='relational noun'.

(4) (Nan R)

a. *Kol* *ch'unumal* *ta'n* *tal* *xuak.*
kol ch'unumal ta'n tal xuak
basket carry by small girl
'Una patoja lleva una canasta.'
'A young girl is carrying a basket.'

b. *Xuak* *mo* *señora?*
xuak mo señora
girl or woman
'Una patoja o una señora?'
'A girl or a woman?'

c. *Qes* *vete'.*
qes **vet-e'**
old DM-FV
'Ya está anciana.'
'She's already old.'

The uses of *vet* in this narrative reflect the descriptive nature of the task. In sentences (3a) and (4c), *vet* is not used as a way of indicating the sequencing of activities, but rather to mark the end of a sequence of talk and also to highlight the conclusion that has been drawn in the negotiative process of interpreting the pictures. The use of *vet* in (4c) functions both as an adverb and as a metatextual marker of the speakers' recapitulation of the negotiated descriptions. This is an important function that will be discussed further in sections 4.3, 4.4, and 4.5 below.

It is while describing pages 12 and 13 of the book that Nan R, Nan Xh, and Nan P discover that the pictures constitute a sequence that tells a story. Together they describe the picture on page 12 of the woman in the barn milking her cow with her cat watching. They describe the woman on the following page pouring water into a bowl for the cat. When asked if it was water or milk, Nan R looks back to the picture on page 12 and then to the liquid being poured on page 13, then covers her mouth and gasps. She excitedly points at page 12 and then at page 13, claps her hands with glee, and explains to Nan Xh and Nan P that the woman is milking the cow on page 12 and then bringing the milk into the house to give it to the cat, and that the pictures tell a story. With this revelation, all three women clap and exclaim.

This discovery changed the women's activity from giving separate descriptions of the pictures on the pages to telling a story in which the activity of one page leads to the activity on the next. Accordingly, it is at this point that we see that the use of *vet* to indicate the sequencing of activities emerges, as in (5a-c) where the story protagonist's cleaning activities are described:

(5) a. *Aa* *ni* *tx'aa* *vet* *aak* *tul* *b'aj.*
aa n-i tx'aa **vet** aak tul b'aj
FOC INCOMP-3S.ERG wash DM person.of.respect inside well
'Está lavando (la parte de) adentro, pues.'
'She is washing inside then.'

b. *Ni* *taq'* *kuu* *aak*
n-i t-aq' kuu aak
INCOMP-EV 3S.ERG-put DIR:downwards person.of.respect

s-t-ul *la* *qal* *vete'.*
stul la q-al **vet**-e'
RN-3S.ERG-inside POT 1P.ERG-say DM-FV
'Ella lo está echando adentro, digamos.'
'She is putting it inside let's say.'

c. *Il* *chan* *ile'* *at* *je'* *vi'* *u* *meexha.*
il chan ile' at je' vi' u meexha
see PART there.it.is exist DIR:upward on.top.of DET table
'Allí está encima de la mesa.'
'See, it is on top of the table.'

The moment of recognition of the sequential nature of the story in the book represents the women's shift from a pre-iconographic to an iconographic interpretation of the printed page (Panofsky 1972).[5] A pre-iconographic interpretation is a recognition of 'pure forms', the association of the visual experience of the drawings with memory of those items. Because many of the items were foreign to the women or did not occur in situations recognizable to them, they required identification before they could become representatives of things or actors within a narrative. *Iconographic* interpretation is the recognition of shared cultural contexts, the shared themes, concepts, and conventions that give the graphic representation a cultural meaning. After it had been identified by the researchers, one of the women labeled the dress on a hanger as a *chik,* the Ixil word for the wrap skirt the women wear, thus putting it into a meaningful context for her partners. It was only at this iconographic stage in the interpretation of the drawings on the printed page that the women could begin to construct a narrative. And it was only with the start of a narrative that sequencing adverbs and discourse markers became relevant contributors to the structuring of the discourse. This explains the paucity of tokens of *vet* in any function in the preceding IUs of the hundred clauses we analyzed from this narrative and, because this sequence occurs close to the end of those hundred clauses, it also accounts for the low number of occurrences of *vet* found in this sample overall.

[5] We are grateful to Michael McDuffie who directed us to this literature.

4.2. OCCURRENCES OF *VET* IN A *FROG, WHERE ARE YOU?* NARRATIVE. *Frog, Where are You?* was the first book that several of the members of the *Grupo de Mujeres por la Paz* had ever held in their hands, and these women were visibly nervous when they were asked to look at the book and tell each other the story of what they saw in it. Many of them initially declined the task because they did not know how to read, but agreed to participate when they were told that the book contained only pictures, not words. The pictures in *Frog, Where are You?* tell the story of a boy and his dog who have captured a frog and put him in a jar. On the first page of the book, it is nighttime and the boy and the dog are lying on the floor in a bedroom, looking at the frog in the jar. When the frog escapes overnight, the boy and the dog set off to find the frog, encountering a number of adventures along the way. The dog falls out of the bedroom window, is chased by bees, and finds a groundhog in a hole. The boy climbs a tree, is frightened by an owl and falls to the ground, falls off a cliff and is carried away on the horns of a large deer. In the end, they find the frog, which has reunited with its family.

Like the drawings in *Pancakes for Breakfast,* those in *Frog, Where are You?* contained cultural items and activities that were unfamiliar to the members of the *Grupo*. The first page was confounding for many of them as some thought that the four-poster bed was an altar, an interpretation that was reinforced by the bell-shaped lamp hanging over the bed. Some of the women thought that a striped tee shirt lying on the floor was a pile of firewood, which would be, for them, an appropriate thing to have on the floor, whereas a tee shirt would not. They thought that the cross-hatching used by the artist to indicate the darkness of nighttime outside the window and the shadow inside the room was barbed wire. Some were annoyed that a dog had been allowed in the house and those who recognized the bed as a bed were concerned that a child was sleeping alone. Only a few recognized the frog as a frog.

The three women who participated in the narrative analyzed here are Nan M, MB, and MM. Nan M is a grandmother, the matriarch of a large family and a successful business-woman who has no formal education. MB and MM are unmarried young women in their late teens, the youngest and two of the most fluently bilingual in Ixil and in Spanish of the members of the *Grupo*. They have both had a few years of formal education.

Frog, Where are You? is clearly a story in which reporting the sequencing of events is important and we had expected that these retellings would require more use of adverbial *vet* to move the story line forward and to move the characters in the story from one activity to another. With the two younger women who had had some experience not only with writing but with drawings, this expectation was fulfilled. MM and MB used the adverbial *vet* to indicate the temporal order of activities, as illustrated in (6a-c), and in these cases, the Ixil women have translated *vet* as the adverb ‘now’:

(6) a. *Koxhle’l* *vet* *naj.*
koxhle’l **vet** naj
POS:lying.down now he
‘Él ya está acostado.’
‘Now a boy is lying down.’

b. *Il* *naj* *ile'*;
il naj ile'
look he there.it.is
'Allí está él;'
'There he is;'

kat *imujluj* *vet* *tib'*.
kat i-mujluj **vet** t-ib'
COMP 3S.ERG-hide now 3S.ERG-RN
'ya se escondió.'
'now he has hidden.'

c. *Ile'* *ta'*;
ile' ta';
there.it.is EMP
'Allí está;'
'There he is;'

ni *sik'in* *vete'*.
n-i sik'in **vet**-e'.
INCOMP-3S.ERG shout now-FV
'ya está gritando.'
'now he is shouting.'

In addition to using *vet* to situate activities in a sequence of activities, there were also several instances where *vet* was used to mark the beginning of a sequence after a set of background observations. In the following example, the first IU (7a) opens the sequence that provides the context for the last IU (7d), where *vet* is used to indicate the end of one activity in the story and the beginning of the next. In the pages prior to this one, an owl has frightened the boy, who then falls from the tree to the ground. At the same time, the dog has disturbed a beehive and is being chased off by the bees as the boy is falling. The boy gets up, climbs up on a rock, and calls his dog. A deer emerges from behind the rock and lifts the boy onto his antlers. At the same moment, the dog comes back and goes behind the rock where the deer is. The dog is found just as a new adventure is beginning.

(7) (MM)

a. *Aa* *kat* *jee* *kat* *ch'u'l* *u* *chee* *ti* *u* *k'ub'e'*.
aa kat jee kat ch'u'l u chee ti u k'ub'-e'
FOC COMP go.up LOC DIR DET dear on DET rock-FV
'Allí sale el venado detrás de una piedra.'
'There the deer comes out from behind a rock.'

b. *Vatz* *naj* *kat* *oon* *kat* *vet* *u* *txooe'.*
vatz naj kat oon kat vet u txoo-e'
front.of he COMP arrive LOC now DET animal-FV
'En frente de él llegó el animal.'
'The animal came in front of him.'

c. *Il* *naj* *ile'* *tul* *vete'.*
il naj ile' t-ul **vet**-e' _
there.it.is he there.it.is 3S.ERG-come DM-FV
'Allí ya viene él.'
'There he comes.'

d. *Il* *tal* *itx'i'* *naj* *ile.*
il t-al i-tx'i' naj ile'
there.it.is small 3S.ERG-dog he there.it.is
'Allí está su perrito (de él).'
'There is his little dog.'

e. *Il* *naj* *ile'* *tul* *vete'.*
il naj ile t-ul **vet**-e'
there.it.is he there.it.is 3S.ERG-come DM-FV
'Allí está él, ya viene.'
'There he is, he's coming.'

Here we have *vet* performing one of the primary functions of discourse markers, the bracketing or sequencing of units of talk (Schiffrin 1987:31). It serves a dual function here as the adverbial meaning of temporal sequencing extends to also fulfill the pragmatic discourse function of signaling to the listener that a new item introduced into the discourse is sequentially connected to the previous one, thus building coherence in the storytelling.

The proficient storytelling of the younger women who easily interpret the things and the activities in the pictures accounts for the relatively large number of occurrences of *vet* in a sequencing function in the one hundred IUs of the Frog narrative. By contrast, most of Nan M's contributions to this *Frog* narrative were made in the absence of an iconographic sense of characters and items participating in a storyline, just as we saw in the *Pancakes for Breakfast* narrative. The majority of her utterances are questions and repetitions of what the two younger women have said, repetitions often made with several intervening utterances from the other two women as she studies the pictures. MM, MB, and Nan M also use *vet* to perform some of the same metatextual functions used by the women who narrated *Pancakes for Breakfast*, that is, to highlight agreement with a previous speaker's contribution or to highlight the achievement of consensus among group members as they are explaining to Nan M what is happening in the pictures.

In (8a-e) below, Nan M is asking again for identification of something in the picture.

(8) a. (Nan M)

Kam	*vet*	*unq'a*	*vi'le'*	*q'i?*
kam	vet	unq'a	vi'le'	q'i
what	DM	DET.PL	those	DM

'¿Qué son estes pues?'
'What are these?'

b. (MB)
Ch'i'mal.
ch'i'mal
pretty
'Son bonitos.'
'They're pretty.'

c. (Nan M)

Kam	*uve'?*
kam	uve'
what	that

'Eso, qué es?'
'What is that?'

d. (MM)
Ch'i'mal.
ch'i'mal
pretty
'Son bonitos.'
'They're pretty.'

e. (MM)

Tan	*ni*	*sik'in*	*vete'.*
tan	n-i	sik'in	**vet**-e'
because	INCOMP-3S.ERG	shout	DM-FV

'Porque está gritando.'
'Because he's shouting.'

f. (MM)
Ni sik'in vet naj ti' vitx'i'e'.

n-i	sik'in	**vet**	naj	ti'	vi-tx'i'e'
INCOMP-3S.ERG	shout	DM	he	for	3S.ERG-dog

'El está gritando por su perro.'
'He is shouting for his dog.'

In (8a), Nan M is asking her partners to explain what is happening in the picture. Immediately following Nan M's question, MB comments on something in the picture: "Pretty." MM repeats her comment, and then responds to Nan M, first saying that the boy is shouting, and then expanding on this by repeating that he is shouting and adding "for his dog." Nan M's question and MM's two answers are both marked with *vet* which here serves the function of aligning one's question with the other's response.

To explain the greater number of total occurrences of *vet* in the Frog telling than in the other texts, we must take into account not only the storytelling genre, but also these three participants and the work they must do to collaborate in the storytelling process. Two of the women understand the task and have very little trouble interpreting the pictures, while the third member is inexperienced with both this particular type of narration behavior and the prop. All three must then work harder at performing the task and assuring that they are all performing it together. Thus, *vet* is required to fill its adverbial sequencing function and to serve the metatextual functions of overtly recognizing each participant's contributions and maintaining each woman's stance as a participant.

4.3. OCCURRENCES OF VET IN A SPEECH. In Sections 4.1 and 4.2, we alluded to the *metatextual* functions of *vet*, those instances in which *vet* is used as an overt comment on the discourse itself. In (3a) and (4c) from *Pancakes* and (8a, d, and e) from *Frog*, *vet* was used as a way of agreeing with or referring back specifically to something said earlier in the telling. Further evidence of the metatextual use of *vet* can be seen in a speech given by one of the members of the *Grupo de Mujeres por la Paz* during a meeting to discuss the possibilities of beginning literacy training among the women. The speaker here, Nan Ja, is telling of a woman in another group who has learned to read and write. The woman struggled, according to Nan Ja, but achieved her literacy through that struggle. Seven of the twelve occurrences of *vet* in the hundred analyzed clauses of the speech are used to emphasize the accomplishments of this woman, as the listeners are urged to be like her. The sequence in (9a-f) provides five of those uses of *vet*.

(9) (Nan Ja)

a.	*Pero*	*ni*	*vil*	*vet*	*vitz'ib'*
	pero	n-i	v-il	**vet**	vi-tz'ib'
	CONJ	INCOMP-EV	1S.ERG-see	DM 3S.ERG	letter

aake'.
aak-e'
person.of.respect-FV
'Pero ya veo la letra de ella.'
'But now I see her writing.'

b. *Cheel il lab' ile'*
cheel il lab' ile'
now look that.thing there.it.is
'Ahora mira eso,'
'Now look at that,'

c. *ni tx'ol vet aaki.*
n-i tx'ol **vet** aak-e'
INCOMP-3S.ERG can DM person.of.respect-FV
'ella ya puede.'
'she can do it now.'

d. *Aal b'a'n vet vitz'ib' aake'.*
aal b'a'n **vet** vi-tz'ib' aak-e'
EMP good DM 3S.ERG-letter person.of.respect-FV
'Ya es buena su letra de ella.'
'Now her writing is good.'

e. *Aan chit aak kat vil vet cheel.*
aan **chit** aak kat v-il **vet** **cheel**
just.now only person.of.respect COMP 1S.ERG-see DM now
'Hasta ahorita que yo la veo.'
'It's just now that I see her (notebook).'

f. *Aan chita' cheel ve ni tx'ol vet aak cheel.*
aan chita' cheel ve' n-i tx'ol **vet** aak cheel
just.now only now REL INCOMP-3S.ERG can DM person.of.respect now
'Ella puede hasta ahorita ya.'
'It's just now that she can do it.'

That *vet* is performing both an adverbial and a discourse function in the sentences in (9) is clear from its patterns of occurrence with temporal adverbs *cheel* and *aan.* In sentence (9a), *vet* alone marks the temporal frame of current ongoing activity, bracketing it as a new activity by its occurrence with the conjunction *pero* 'but'. In (9b), the adverb *cheel* occurs in the exhortation clause *Cheel il lab'ile'* 'Now look at that,' while *vet* continues as the temporal marker of the writing activity in (9c) and (9d). In (9e), Nan Ja emphasizes the evidence of the woman's recent accomplishment through the use of the temporal adverb *aan* 'just now' with the particle *chit* 'only', which serves here as an emphatic. The verb *vil* 'I see' is the focus here, with further emphasis of the recentness of her achievement given by use of the collocation *vet cheel* 'now'. This form is used again in (9f) as modification within the nuclear sentence. In this instance the speaker focuses the subject by inserting it between *vet* and *cheel*. Throughout this sequence, *vet* does not simply function as an

adverb indicating temporality. In (9a), it also marks an evaluative statement, strengthening the exhortation to the women of the *Grupo*, making a metatextual comment to highlight her point and request agreement.

In (9d), *vet* highlights how good the woman's writing is. In (9e) the speaker asserts that she, herself, has seen the notebook, has seen the improvement, and inserts herself as a witness, thereby asserting her authority to comment on the virtues and outcomes of the writer's efforts. The conclusion in (9f) returns the focus to the woman and her accomplishment. *Vet* has the further metatextual function of inviting listeners to participate in the evaluation of the benefits of writing and to aspire to those benefits for themselves. This use of *vet* is similar to the use in (4) from the *Pancakes* narrative in that in both cases *vet* is used to reinforce not only the proposition itself but, more importantly, the speaker's testimony about the state as something that is already the product of agreement.

4.4. OCCURRENCES OF *VET* IN A MEETING DISCUSSION. The last genre of discourse to be examined here is group discussion at a meeting called by the *Grupo* directive board to explain and discuss the Master's thesis García was planning to write on the stories told by members of the group (García 2005). Although there were very few instances of *vet* in the 100 IUs extracted for review from the recording of the meeting, the examples we focus on here come in combination with another frequent discourse marker in Ixil Maya, *la qale'*. We thus begin here with a discussion of *la qale'* before presenting the examples of *vet* from the meeting transcript.

4.4.1. THE DISCOURSE MARKER *LA QALE'*. *La qalchaj* or *la qale'*, is translated by the women as 'digamos' in Spanish and by the researchers as 'let's say' in English. It consists of the potential aspect marker *la* and the verb root for 'say', *-al*, plus the first person plural ergative prefix *q-* and the final vowel suffix *-e*.[6]

La qale' has several functions that are extensions of its literal meaning of 'digamos' or 'let's say'. One of these functions is to pose hypothetical situations, as in (10).

(10) (Nan Mt)

Kam	*tatin*	*u*	*gruupoe'*	*la*	*qale',*
kam	t-atin	u	gruupo-e'	**la**	**q-al-e'**
how	3S.ERG-way.of.being	DET	group-FV	POT	1P.ERG-say-FV

'Cómo es el grupo, digamos,'

'How the group is, let's say,'

[6] Ayres (1991: 162) describes the suffix *-e'* as a 'final suffix' occurring only in phrase-final position and having several functions, among them the marking of an intransitive verb when it occurs in sentence-final position. Because it occurs on many non-verbal words in the Nebaj dialect of Ixil, we are currently analyzing this phrase-final *-e'* as a phonological rather than a morphological feature and have, therefore, labeled it as a 'final vowel'.

la	*tale'*	*chi*	*ni*	*tal*	*ixoj*.
la	t-al-e'	chi	n-i	t-al	ixoj
POT	3S.ERG-say-FV	they.say	INCOMP-EV	3S.ERG-say	she

'ella va a explicar lo que ella está diciendo.'
'she's going to explain what she's saying.'

In (11), we see *la qal* participating in a fixed phrase meaning 'for example'.

(11) (Nan Mt)

Porque	*la*	*qal*	*b'en*	*sti'*	*kat*	*ich'oti*	*ixoj*.
porque	**la**	**q-al**	**b'en**	**s-t-i'**	kat	i-ch'oti	ixoj
because	POT	1P.ERG-say	DIR	RN-3S.ERG-RN	COMP	3S.ERG-ask	she

kam	*ixe't*	*u*	*gruupoe'*.
kam	i-xe't	u	gruupo-e'
how	3S.ERG-begin	DET	group-FV

'Porque digamos preguntó ella que cómo empezó el grupo.'
'Because let's say that she asked how the group began.'

It is also used to unify group members in their experiences and their evaluations of those experiences, as in (12):

(12) (Nan Mt)

a.

Kam	*sti'*	*nuk'el*	*kat*	*qib'*	*la*	*qale',*
kam	s-t-i'	nuk'el	kat	q-ib'	**la**	**q-al-e'**
why	RN-3S.ERG-RN	unite	LOC	1P.ERG-RN	POT	1P.ERG-say-FV

'¿Por qué estamos reunidos, digamos?'
'Why are we united, let's say,'

b.

tan	*ti*	*u*	*txumleb'ale'*
tan	ti	u	txumb'leb'al-e'
for	to	DET	great.sadness-FV

'solo por la gran tristeza'
'just because of the great sadness'

c.

at	*tu*	*vas*	*tename'*	*la*	*qale'*.
at	tu	vas	tenam-e'	**la**	**q-al-e'**
exist	in	[PART]	town-FV	POT	1P.ERG-say-FV

'que hay en el pueblo, digamos.'
'that there is in town, let's say.'

And *la qale',* or its variant *la qalaj,* is also used to encourage interlocutors to participate in the evaluation of current situations, inviting them to share the speaker's stance rather than presenting that stance as fact, as in (13a-b):

(13) (Jac)

a. *Ye'* *ni* *qab'i* *ve* *kam* *ni*
ye' n-i q-ab'i ve kam n-i
NEG INCOMP-EV 1P.ERG-listen CPZ what INCOMP-EV

talaxe' *ve* *la* *qalaj.*
t-al-ax-e' ve la **q-al-aj**
3P.ERG-say-PAS-FV CPZ POT 1P.ERG-say-FV
'Ya no escuchamos lo que están hablando, digamos.'
'We don't hear what they're saying, let's say.'

b. *Pes* *tzitzi* *ta'* *kat* *tz'ejxi* *kat* *o'* *la* *qale'.*
Pes tzitzi ta' kat tz'ej-x-i kat o' **la** **q-al-e'**
well there EMP COMP lose-PAS-FV LOC 1P.ABS POT 1P.ERG-say-FV
'Pues allí es donde ya nos perdimos, digamos.'
'Well that's where we've lost our way, let's say.'

While the many functions of *la qale'* as used by the women could be further demonstrated, the preceding examples suffice to indicate that it is used to elicit participation from interlocutors in a statement of purpose by giving examples of desirable behaviors or outcomes or an evaluation of a situation. The relationship and co-occurrence of *la qale'* and *vet* as discourse markers will be examined in 4.4.2 below.

4.4.2. OCCURRENCES OF *VET* IN THE THESIS MEETING. The hundred clauses of the *Thesis Meeting* recording selected for analysis are highly representative of the rest of the meeting discussion for their remarkable lack of usage of *vet* in any of its functions. During this meeting, Nan Mt, a member of the *Grupo* board of directors explains what a Master's degree thesis is, what author García's thesis will be about, and why the members of the *Grupo* have become the subject of this work. No sequencing of activities is involved in the straightforward explanation that a thesis is like a book and that this book will be about their activities during and after the war. The following translation into English from the Ixil contains (10) above and the description of their task:

(14) (Nan Mt)

‘We’re going to write a book of information, she says, about how the *Grupo* began, why the *Grupo* was formed. What the *Grupo* is like, let’s say. She’s going to explain what she’s saying (what she means). And she says that you can speak in Ixil because “Ixil interests me,” she says. We are not going to think that our language has no value, but instead that it is good that we say it in Ixil. How the *Grupo* is, let’s say, because that is what people want, so that people know how the *Grupo* is here. Why we are united, let’s say, is what we’re going to teach them. They’re going to give her a little money and that is how she comes so often.’

While *vet* as a sequencing discourse marker is not found in these lines from the thesis meeting, *vet* does appear with metatextual function, in the lines that immediately follow, (15a-c)**:**

(15) (Nan Mt)

a. *La* *kub’an* *vet* *tuch* *unq’a* *u’uje’.*
la ku-b’an **vet** tuch unq’a u’uj-e’
POT 1P.ERG-do DM prepared DET.PL paper-FV
‘Los vamos a preparar los papeles.’
‘We’re going to prepare the papers.’

b. *Tu* *yolb’al* *la* *qal* *kat* *vet* *lab’.*
tu yol-b’al la **q-al** kat **vet** lab’
in word-INST POT 1P.ERG-say LOC DM it
‘En Ixil lo vamos a decir.’
‘We’re going to tell it in Ixil.’

c. *Li* *tal* *vete’* *kam* *ixe’t* *u* *gruupoe’,*
li t-al **vet**-e’ kam i-xe’t u gruupo-e’
POT 3.ERG-say DM-FV how 3S.ERG-begin DET group-FV

la *qal* *vete’,*
la **q-al** **vet**-e’
POT 1P.ERG-say DM-FV
‘van a decir cómo empezó el grupo, digamos,’
‘it’s going to say how the Group began, let’s say,’

kam *ni* *taq’oma* *u* *gruupoe’.*
kam n-i t-aq’oma u gruupo-e’.
what INCOMP-3S.ERG 3S.ERG-work DET group-FV
‘qué trabajo hace el grupo.’
‘what work the group does.’

In (15), Nan Mt gives the description and explanation as a frame for the project the group is about to embark upon, framing the shared commitment and collective action expected of the women (Benford and Snow 2000, Goffman 1959). Such framing acts are "intended to mobilize potential adherents and constituents, to garner bystander support, and to demobilize antagonists" (Snow and Benford 1988:198). The resulting collective action frame is presented through a discussion of shared activities and memories that, as Gamson puts it, "are not merely aggregations of individual attitudes and perceptions but also the outcome of negotiating shared meaning" (Gamson 1992:111). Similarly, (15b-c) shows *vet* used to frame discourse as communal, and it is here that we see *vet* co-occurring with *la qal/la qale'*.

The use of *vet* here has the same metatextual force as its usage in the previous examples, that is, *vet* highlights the proposition as one that is the result of collective agreement as in (4c) and (8a-f) or as one that the speaker is posing as likely to be received with collective agreement as in (13a-b). The use of *la qal vete'* in (15c) exhorts the women as a group to participate through the imposition of a logical sequence of activities that will end with their agreement with the negotiation. The use of *vet* in this instance has the same force as its use in the description of the woman's writing; that is, it highlights the description of something already accomplished by means of group negotiation, and thus serves as a marker of group consensus.

5. CONCLUSIONS. In the analysis of the four interactions we can see that *vet* has both adverbial and pragmatic functions. As an adverb it allows the structural sequencing of activities although it seems to be temporally fluid, as it occurs with the completive aspect, with the potential and incompletive aspects, and with states-of-being that are uninflected for aspect. It has the significance of 'now', 'then', 'already', and 'still'. This temporal fluidity points to the pragmatic function of directing the listeners' attention to the relationship between activities, allowing for its use in showing the relationship of activities to the external as well as the internal discourse context.

This range of meaning, from adverbial to pragmatic, is in line with Traugott's (1997) hypothesis about the development of discourse markers. She proposes the following grammaticalization cline for discourse markers:

Adverbial > Sentence Adverbial > Discourse Particle (including DM)

Grammaticalization refers to those "linguistic changes through which a lexical item in certain uses becomes a grammatical item, or through which a grammatical item becomes more grammatical" (Hopper and Traugott 1993:2). Along with grammaticalization comes a "parallel semantic change" in which a concrete item becomes more abstract and, with that abstraction, comes a bleeding of the concrete meaning (Traugott 1997:2). We hypothesize that the adverbial use of *vet* as an indicator of the sequential relationships between activities, has developed into the more pragmatic and less semantically concrete function of drawing the listeners into complicity with the speaker by emphasizing prior consensus on a particular point, and further, indicating a consensus that is only assumed to prevail.

The absence of *vet* in the pre-iconographic attempts at descriptions in the *Pancakes for Breakfast* text reveals the importance of this adverbial function. Once the women have

decoded the pictures and recognized their sequential, storytelling nature, their narration is more story-like, thus requiring temporal connections through adverbial uses of *vet.* The important pragmatic functions of *vet* are revealed when it is used to signal acceptance of the negotiation of description in the picture book retellings. And just as the women signal descriptions that have been agreed upon by using *vet*, they also use *vet* to express a particular evaluative stance, that is, that the expression is judged by the speaker as truthful and is likely to be received by the group with agreement.

As we have seen with *vet*, various stages of change can co-exist in the grammaticalization process. The period of overlap in which older and newer meanings co-exist is referred to as 'layering' (Hopper 1991, Hopper and Traugott 1993), with the original meaning remaining in use separately or concurrently with the grammaticalized meaning(s). Throughout our analysis, we have seen instances of *vet* serving an adverbial function of marking the sequential relationships between clauses while at the same time performing the pragmatic function of drawing the listeners into complicity with the speaker by accepting the results of negotiated meanings.

We began this paper with a summary of the women's social and historical background because we believe that this kind of context is crucial to the development and use of discourse markers like *vet.* We also began with hypotheses about the nature and number of the usage of *vet* as factors of genre. What we see throughout the analysis is the importance of the effects of the individual identities of the participants in each discourse outweighing the effects of the genre of the discourse. One of the most relevant facts of our fieldwork in Nebaj is the varying levels of literacy skills among the women and the correspondence of those skills with the differentials in experience the women have had with printed materials. This fact has determined how transcription and elicitation sessions will be arranged and now it is clear that it is also a factor in shaping how, and how frequently, the discourse marker *vet* is used by the women in their speech.

But each of the women with whom we work also has an identity as a member of the *Grupo de Mujeres por la Paz.* The external context surrounding all four of these texts have crucially to do with the women's membership in a group formed to support each other in their struggles to recover from the wartime losses of family, land, and community that they all experienced. In the six years we have been working with the *Grupo de Mujeres por la Paz,* we have seen them work *as a group*, making plans and decisions as a group, caring for the sick among them as a group, learning to read and write as a group (in their initial discussions about the literacy classes, the question about participation of the older women arose and they agreed that *all* of the women would participate together with the younger women helping the older women, as necessary). No significant decision is made without a general meeting that includes input from all who wish to speak. It is important to the success of any of their enterprises that they negotiate meanings before they make decisions, and it can be hypothesized that these negotiations contribute to construction and maintenance of their solidarity. That this motivation will be expressed and recreated in their discourse practices and reflected grammatically is to be expected.

References

Ayres, Glenn. 1991. *La grammatica Ixil*. Antigua, Guatemala: Centro de Investigaciones Regionales de Mexoamérica.

Benford, Robert D. & David Snow. 2000. Framing processes and social movements: An overview and assessment. *Annual Review of Sociology* 26. 611-639.

Chafe, Wallace. 1994. *Discourse, consciousness and time.* Chicago: The University of Chicago Press.

Crismore, Avon, Raija Markkanen & Margaret S. Steffensen. 1993. Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication* 10(1). 39-71.

diPaolo, Tomie. 1978. *Pancakes for breakfast*. New York: Voyager Books.

England, Nora C. 2007. Marking aspect and mood and inferring time in Mam (Mayan). Paper presented at the Berkeley Linguistics Society. University of California, Berkeley.

Falla, Ricardo S.J. 1998. *A story of a great love: Life with the Guatemalan "communities of population in resistance"*. Washington DC: EPICA.

Ford, Cecelia E., Barbara A. Fox & Sandra A. Thompson. 2002. *The language of turn and sequence*. New York: Oxford University Press.

Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31. 931-952.

Gamson, William A. 1992. *Talking politics*. New York: Cambridge University Press.

García, María Luz. 2005. Woven words: Constructing cohesion in Ixil Mayan women's discourses. Austin, TX: University of Texas MA thesis.

Goffman, Erving. 1959. *Frame analysis: An essay on the organization of the experience*. New York: Harper Colophon.

Hanks, William. 1987. Discourse genres in a theory of practice. *American Ethnologist* 14(4). 668-692.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, 17-35. Amsterdam & Philadelphia: John Benjamins.

Hopper, Paul J. & Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Johnstone, Barbara. 1990. *Stories, community, and place: Narratives for middle America*. Bloomington: Indiana University Press.

Johnstone, Barbara. 2002. *Discourse analysis*. Oxford: Blackwell.

Manz, Beatriz. 1988. *Refugees of a hidden war: The aftermath of counterinsurgency in Guatemala*. Albany: State University of New York Press.

Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Books.

Panofsky, Erwin. 1972 (1939). *Studies in iconology: Humanistic themes in the art of the Renaissance*. Boulder: Westview Press.

Redeker, Gisela. 1991. Review article of "Linguistic markers of discourse structure". *Linguistics* 29(6). 1139-1172.

Sanford, Victoria. 2003. *Buried secrets: Truth and human rights in Guatemala*. New York: Palgrave Macmillan.

Schiffrin, Deborah. 1987. *Discourse markers*. New York: Cambridge University Press.

Schiffrin, Deborah. 2001. Discourse markers: Language, meaning, and context. In Deborah Schiffrin, Deborah Tannen & Heidi Hamilton (eds.), *Handbook of discourse analysis*, 54-75. Oxford: Blackwell-Wiley.

Snow, David & Robert D. Benford. 1988. Ideology, frame resonance, and participant mobilization. *International Social Movement Research* 1. 197-218.

Traugott, Elizabeth Closs. 1997. The role of the development of discourse markers in a theory of grammaticalization. Paper presented at the International Conference on Historical Linguistics XII. Manchester.

Jule Gómez de García
jmgarcia@csusm.edu

Melissa Axelrod
axelrod@unm.edu

María Luz García
mluzgarcia@yahoo.com

3

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 33-56
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4450

Classifying clitics in Sm'algyax: Approaching theory from the field

Jean Mulder and Holly Sellers

University of Melbourne

Sm'algyax (British Columbia and Alaska) is a highly ergative VAO/VS language with an uncommonly wide range of clitics. This chapter has the two-fold function of demonstrating how Anderson's (2005) constraint-based analysis of clitics gives insight into the complex behavior of Sm'algyax clitics, and how the clitics themselves afford empirical means of testing such a theory. The Sm'algyax data are drawn from both field research and published texts, reflecting a community-based approach to language documentation that has evolved through a long-term, collaborative relationship with the Tsimshian (Sm'algyax) communities. Building on Stebbin's (2003) definitions of intermediate word classes in Sm'algyax and Anderson's Optimality Theoretical approach, we determine that in terms of their varying phonological dependence, Sm'algyax clitics include internal, phonological word, and affixal clitics. The existence of affixal clitics in Sm'algyax, however, calls into question the viability of the Strict Layer Hypothesis (Selkirk 1984) as inviolable rules when describing clitics. Furthermore, Sm'algyax provides strong evidence that the direction of clitic attachment is more clitic specific than language specific. In characterising the behaviour of Sm'algyax clitics, we find that not only does linguistic theory help sharpen our understanding of the fieldwork data, but also that field linguistics has consequences for linguistic theory.

1. INTRODUCTION. Sm'algyax (Coast Tsimshian), an endangered member of the Tsimshianic language family, is a highly ergative, mildly polysynthetic VAO/VS language spoken by the Tsimshian people in the northernmost area of coastal British Columbia, Canada and since the late nineteenth century in the neighboring region in Alaska. In Canada there are seven Tsimshian bands, with a total of approximately 7,700 band members. While there is no accurate survey of the total number of fluent speakers, Sm'algyax is highly endangered with complete speaking command mostly restricted to older age groups. The extent of the shift toward English correlates not only with age but also with the population size and degree of geographic isolation of the community. However, over the last 30 years the Tsimshian people have been actively working to preserve and revitalize their traditional language through a range of activities involving school-based language programs, curriculum development, literacy training, native language teacher training and language documentation. Linguistic description of Sm'algyax began with fieldwork by Franz Boas (e.g., 1911, 1912) and then most significantly continued from 1969 with the efforts of John

Dunn (e.g., 1978, 1979a), as well as the first author's own fieldwork begun in 1979 (e.g., 1994), and Tonya Stebbins' fieldwork from 1996 (e.g. 1999, 2001a).[1]

An area that has been challenging for field linguists has been characterizing the behavior of the uncommonly wide range of clitics in Sm'algyax. Both grammatical and lexical classes of clitics occur and include a large number of proclitics and enclitics, a flexiclitic (attaching variably as a proclitic or an enclitic) and a circumclitic. The clitics vary most notably in terms of their phonological dependence and their direction of attachment. To illustrate, in (1) the 'predicative connective' enclitic *=sga* carries grammatical-semantic information about the following NP, *nliit* 'his fur', but is prosodically attached to its left, whereas the possessive proclitic *n=* also occurs at the left edge of the NP, but attaches to the right.[2]

(1) *Ada gyik loogaksga nliit.*
Ada= gyik loogak[=sga n=lii-t]
and= again wet[=CN.PRED POSS=fur-3POSS]
'And his fur was wet again.' (Mulder 1994:195)

This chapter has the two-fold function of demonstrating how Anderson's (2005) constraint-based analysis of clitics gives insight into the complex behavior of Sm'algyax clitics, and how the Sm'algyax clitics themselves afford empirical means of testing such a theory. This is a case where not only does linguistic theory help sharpen our understanding of fieldwork data, but also where field linguistics has consequences for linguistic theory.

The data for this chapter are drawn from texts from field research by the first author, some published (e.g., Mulder 1994), but many unpublished; texts and examples collected from speakers in the field by Boas (1911, 1912), Dunn (1978, 1979a, 1995) and Stebbins (1999, 2001a, 2003); and published texts, written and edited by Sm'algyax speakers (e.g., Bolton et al. 1984; Reeves 1992). With the exception of Boas' materials, the texts specifically drawn on for this study have all been edited by a collective of fluent Sm'algyax speakers who are also proficient writers of their language. The understanding of the data presented here is framed by the five years the first author initially spent full time in the field doing language documentation and developing the Sm'algyax Language Studies Program in School District No. 52 (British Columbia, Canada). The approach to the data and the choice of texts that underpins this chapter reflects a community-based orientation to fieldwork and language documentation that has evolved through a long-term, collaborative relationship by the first author with a number of people in the Tsimshian communities. The motivation for approaching fieldwork and theoretical linguistic analysis in this way is to

---

[1] For a detailed discussion of linguistic material on Sm'algyax see Mulder (1994) and Stebbins (1999).

[2] The following abbreviations are used for glosses: A='most agent-like argument of a transitive verb'; CN.ATTR='attributive connective'; CN.PRED='predicative connective'; CN.PREP='prepositional connective'; EVID='evidential'; FUT='future tense'; NEG='negator'; O='most patient-like argument of a transitive verb'; PAST='past tense'; PL='plural'; POSS='possessive'; PREP='preposition'; S='argument of an intransitive verb'; SG='singular'; SUBORD='subordinator'.

work toward enabling knowledge of the language to be constructed not only *for* and *with*, but also *by* community-members. Indeed, the body of texts and curriculum that has and is continuing to be co-produced by fluent speakers and proficient writers of Sm'algyax supports such an approach.

The aim of this chapter is not to give a comprehensive account of all the sets of clitics occurring in Sm'algyax, but to provide a descriptive account and theoretical analysis of a subset representing the range of clitics found in this language. The first major part of the chapter begins with a brief presentation of the six sets of Sm'algyax clitics that are considered here (Section 2.1). Building on Stebbins' (2003) definitions of intermediate word classes in Sm'algyax, we then outline language-specific criteria for distinguishing the varying phonological dependence of Sm'algyax clitics (Section 2.2). It is proposed that Sm'algyax clitics either attach as part of the same phonological word (PWd) as their host (termed an Internal Clitic), have independent status as a PWd and form a phonological phrase (PPh) with their host (a PWd clitic), or even join their host to form a recursive PWd structure (an Affixal Clitic). Turning to the morphosyntactic dimension of cliticization, we characterize the position of Sm'algyax clitics in terms of their association with a syntactic domain, their location within this domain and their direction of attachment to their host (Section 2.3). This part of the chapter concludes with a summary that situates Sm'algyax clitic features within the broader typology of clitics (Section 2.4).

In the second major part of the chapter we focus on the implications of the Sm'algyax data for the theory of clitics. After a brief outline of Anderson's (2005) Optimality Theoretic approach to cliticization (Section 3.1), we consider the evidence from Sm'algyax that the direction of clitic attachment is clitic-specific rather than language-specific (Section 3.2). Following this we demonstrate that unlike K^{w}ak^{w}ala, Sm'algyax allows nominals with left-edge clitics and leftward clitic attachment as clause initial topics, thereby challenging the view that the lack of a host precludes clause initial topics (Section 3.3). In the final section we return to the Affixal Clitics in Sm'algyax and discuss how their presence adds further empirical support to Anderson's constraint-based approach to the theory of clitics.

2. CLASSIFYING SM'ALGYAX CLITICS.

2.1 A SELECTION OF SM'ALGYAX CLITICS. While there are numerous clitics in Sm'algyax, in this chapter we consider only the following sets, which are representative of the range of clitic properties in this language:

- predicative connectives
- attributive connectives
- the possessive clitic
- modifier and locative clitics
- clause initial clitics
- subjective dependent pronouns

Here we introduce each set to give a feel for the data and in Sections 2.2 and 2.3 return to a more detailed discussion of their clitic properties. The first two sets, the predicative and attributive clitics, are part of a larger system of what Boas (1911) originally termed 'connectives' in that they mark the functional relationship of the phrase to the larger construction.[3] In addition, most of the sets of connectives also mark several other grammatical and semantic distinctions, such as the predicative connectives, which can mark whether the referent of the noun phrase is a common or proper noun, what its deictic status is relative to the speech act, what its grammatical relationship is to the verb and whether it occurs in an indicative or subjunctive clause. As illustrated in (2) (as well as in (1) above), the predicative connective *=sga* occurs at the left edge of the noun phrase and attaches as an enclitic to the preceding word.

(2) *Ada ła dm dzaksga łgu gwe'am awta.*
ada= ła= dm= dzak[=sga łgu= [gwe'a=m] awta]
and= near= FUT= die[=CN.PRED little= [poor=CN.ATTR] porcupine]
'And poor little porcupine was about to die.' (Mulder 1994:201)

In contrast to the predicative connectives, there are only two attributive connectives: *=a* and *=m*. As is shown in (2) above, an attributive connective occurs at the right edge of the modifier phrase and attaches to the left (see (3) below for an example with *=a* 'CN.ATTR'). The possessive clitic *na=/n=/na*, which refers to alienable possession, occurs at the left edge of the possessed noun phrase and attaches to the right as *na=* in (3) (and as *n=* in (1) above):

(3) *Na'yeen na'wiiama galmt'u'utsgu.*
na'yeen [na='wi=[aam=a] galmt'u'utsk-u]
rust.spot [POSS=big=[good=CN.ATTR] pot-1SG.POSS]
'My nice big pot has a rust spot.' (Stebbins 2003:404)

The possessive clitic can also be varied in its orthographic representation, sometimes being written as a separate word, as in (4):

(4) *Dm gwildm luwansm a na xsoosm.*
dm= gwildm lu=wan-sm a [na= xsoo-sm]
FUT= get.ready in=sit.PL-2PL.S PREP [POSS= canoe-2PL.POSS]
'You must be sitting ready in your canoes.' (Reeves 1992:16)

[3] Stebbins (2003) instead uses the term 'dependency markers'. However, we continue to follow the tradition in the Tsimshianic literature which uses 'connectives'.

The next set, the modifier and locative clitics, are lexical clitics.[4] These clitics, which modify nouns and verbs, do not serve any grammatical purpose and instead either provide extra descriptive information (in the case of the modifiers) or give information about location (in the case of the locatives) (Stebbins 1999:189).[5] This is shown in (5), where the locative lexical proclitic *dzaga=* 'across' occurs with the verbs *yaa* 'walk.SG' and *goo* 'go.to' while *nsmlaxyuuptga* 'his country' begins with the possessive *n=* followed by two lexical clitics, *sm=* 'real' and *lax=* 'top', all attaching as proclitics to the noun *yuup* 'earth':[6]

(5) *Ada* *wil* *dzagayaasga* *awtaga* *adat*
ada= wil= dzaga=yaa=sga awta=ga ada=t
and= then= across=walk.SG=CN.PRED porcupine=EVID and=3A

dzagagoo *nsmlaxyuuptga.*
dzaga=goo n=sm=lax=yuup-t=ga
across=go.to POSS=real=top=earth-3POSS=EVID
'And then porcupine walked across and went across to his country.' (Mulder 1994:189)

In addition to being lexical, the modifier and locative clitics are also atypical in their phonological shape. Unlike the three sets of clitics we have considered so far, which are either non-syllabic or monosyllabic, many of the modifier and locative clitics are disyllabic (as is *dzaga* 'across' in (5)).

The clause initial clitics are grammatical forms which include clausal conjunctions, subordinators, negators, tense/aspect markers and discourse markers. They occur on the left edge of the clause, usually in the order listed, and like the modifier and locative clitics many are disyllabic and a few are even trisyllabic. For example, in (2) the clause begins with the clausal conjunction *ada* 'and', followed by the tense/aspect marker *ła dm* 'near FUT', whereas in (5) the first clause begins with the clausal conjunction *ada wil* 'and then' and the second clause begins with the clausal conjunction *ada* 'and'. The clause initial clitics are prosodically dependent on the first lexical word of the clause which is either an adverb (as in (1)) or a predicate (which is commonly a verb, as in (2) and both clauses in (5)).

Now the clause initial clitics have consistently been treated in the Tsimshianic literature as a set since they 'huddle together' within the same accentual—that is, phonological—phrase. However, when we look at the syntactic structure of Sm'algyax clauses, it

---

[4] As Stebbins (1999, 2003) notes, there are derivational prefixes which perform a similar descriptive function as the lexical clitics, the crucial difference being however, that the prefixes change the meaning of the stem they attach to whereas the clitics do not. Evidence for these modifiers and locatives being classed as clitics, rather than as words, is that they occur in a fixed order and are prosodically dependent on the following word.

[5] A complete list of forms along with a detailed discussion of their distribution is given in Stebbins (1999).

[6] As *=ga* illustrates in (5), there is a small set of clause-final evidential enclitics in Sm'algyax.

appears that while the clausal conjunctions, subordinators and negators are positioned with respect to the left edge of the clause, the tense/aspect and discourse markers are positioned with respect to the left edge of the phrase formed by the predicate and its modifiers rather than the clause. Although fully presenting the evidence for this claim would take us too far afield, we do offer three examples in support of this analysis. In (6) a discourse marker precedes the verb:

(6) *'Ap libagaythaw gyet da awaan.*
'ap= libagayt=haw gyet da awaan
really mixed.up=say person PREP over.there
'That person is not making sense (talking crazily).' (Stebbins 1999:194)

Here the discourse marker *'ap*= 'really', gives emphasis to the predicate and along with the modifier proclitic *libagayt*= 'mixed up' is prosodically dependent on the verb *haw* 'say'. This ordering, which is quite fixed, suggests that the discourse marker *'ap*= 'really' and the modifier clitic orient to the head of a 'predicate phrase'. In (7) and (8) the S is a clause initial noun phrase topic:

(7) *'Nüüyu nah algyagada Sm'algyax.*
['nüüyu] nah= algyax-t[=a Sm'algyax]
[1SG.S] PAST speak-3S[=CN.PREP Sm'algyax]
'I was the one who was speaking in Sm'algyax.' (Mulder 1994:65)

(8) *Da wila k'aba hagwil yaatga.*
da= wila= k'aba hagwil= yaa-t=ga
and= then= little.one slowly= walk.SG-3S=EVID
'And then the little one walked along slowly.' (Mulder 1994:202)

As we see in (7), the topic *'nüüyu* '1SG' comes before the tense/aspect marker *nah* 'PAST' and the verb while in (8) the topic *k'aba* 'little one' comes after the clausal conjunctions *da wila* 'and then' and before the modifier clitic and verb. The ordering in these examples, which again is quite fixed, suggests that the position for clause initial topics is after any clausal conjunctions, subordinators and negators and before any tense/aspect markers and discourse markers of the predicate phrase.

The clause initial clitics are clearly not prototypical in a number of respects, a matter we return to in the following sections. For the moment the focus is the distinction between clause initial clitics which are positioned with respect to the left edge of the clause—the clausal conjunctions, subordinators and negators—and those which are positioned with respect to the left edge of the predicate phrase—the tense/aspect markers and discourse markers — as this is relevant for the final set of clitics we consider here.

In Sm'algyax there are three sets of dependent pronouns, which, following the terminology of Boas (1911) and Dunn (1979b), are the subjective, objective and definite objective dependent pronouns. The choice of dependent pronoun for the A, S and O roles in a clause is conditioned by the tense/aspect and mood of the clause, as well as the person

or semantic nature of the A and O relative to each other. While the objective and definite objective dependent pronouns are suffixes to the verb, the subjective pronouns are clitics (Stebbins 2003:398).[7]

TABLE 1. Subjective dependent pronouns

	SG	PL
1	*n*=	*dp*=
2	*m*=/=*m*	*m*=…=*sm*
3	=*t*	

In terms of position, the subjective dependent pronouns are probably best analyzed as being located with respect to the leftmost element of the predicate phrase—namely, the tense/aspect marker. With a tense/aspect marker, their preferred host, the first person pronouns *n*= '1SG' and *dp*= '1PL', and the second person singular pronoun *m*= '2SG' (as in (9)) are proclitics, while the second person plural pronoun is a circumclitic *m*=...=*sm* '2PL' (as in (10)), and the third person pronoun =*t* '3' is an enclitic (as in (5) above).[8]

(9) *Aam mdm dibaaltga, k'anayis, dm*
Aam m=dm= di=baal-t=ga k̲'anayis, dm=
good 2SG.A=FUT= with=try-3O=EVID friend FUT=

ganaamnt.
gan=aam-n=t
means.of=good-2SG.S= EVID
'It would be good if you tried it too, my friend, you will feel good.' (Mulder 1994:163)

[7] As the subjective pronouns are used exclusively to mark the ergative (i.e. the A role), Stebbins (2003) refers to them as ergative dependent pronouns. However, we continue to follow the tradition in the Tsimshianic literature and use 'subjective dependent pronoun'.

[8] Note that *m*=...=*sm* '2PL' cannot be analyzed as being two separate elements since =*sm* cannot occur on its own and if *m*= occurs on its own it is interpreted as the singular form, *m*= '2SG'.

(10) *Ałga* *mdmsm* *gaba* *ksgoogm* *googa*
ałga= m=dm=sm gap[=a [ksgoox=m] goo[=ga
NEG= 2.A=FUT=2.PL.A eat[=CN.PRED [first=CN.ATTR] thing[= CN.PREP

dmt *gyiinsm.*
dm=t gyiin-sm]]
FUT=3A share.food-2O]]
'Do not eat (the food) they bring you first.' (Stebbins 2003:401)

If there is no preferred tense/aspect clitic host, then a subjective dependent pronoun can stand by itself by being syllabic (as in (11)), or with an epenthetic vowel (as in (12)).[9] In this situation the subjective dependent pronoun, like any clause initial clitic, is prosodically dependent on the first lexical word of the clause.

(11) *Ada* *wil* *m* *wayu.*
ada= wil= m= way-u
and= then= 2SG.A= find-1SG.O
'Then you (SG) found me.' (Boas 1911: 384)

(12) *Gyiloo* *ma* *łaałaga* *gałaxsn.*
gyiloo= ma= łaałax=a ga-łaxs-n
don't= 2SG.A= bite.PL[=CN.PRED PL-fingernails-2POSS]
'Don't bite your fingernails.' (Stebbins 2003:403)

Alternately, the second person singular and the third person subjective dependent pronouns can attach as enclitics to a neighboring clause initial clitic. Thus, the second person singular pronoun is a flexiclitic, attaching variably as a proclitic *m*= '2SG' as in (9), or an enclitic =*m* '2SG' as in (13) where it cliticizes to *gyiloo* 'don't', a negator (see also (5) above where in the second clause =*t* '3' attaches as an enclitic to the clausal conjunction *ada* 'and').

(13) *Gyiloom* *suwanooyu.*
gyiloo=m suwanoo-u
don't=2SG.A bother-1SG.O
'Don't bother me.' (Stebbins 2003:403)

As these examples illustrate, the behavior of the subjective dependent pronouns is anomalous in several respects; first, the pronouns vary across the set in terms of their di-

[9] Note that the first and second person pronouns occur independently much more readily than the third person pronoun (Stebbins 2003:403).

rection of attachment; second, they cliticize to another clitic[10]—namely, one of the clause initial clitics—although they do not form a syntactic or semantic constituent with their host; and third, their preferred host is a tense/aspect marker, but in the absence of one they can attach to another clause initial clitic or to the first lexical word of the clause. These anomalous properties are considered in more detail in the next two sections.

2.2 DISTINGUISHING CLITICS IN SM'ALGYAX. Sm'algyax clitics, as Stebbins (2003) demonstrates, can be positively defined as separate from words and affixes and as dividing into distinct classes. Stebbins concludes that these clitics all 'show some phonological or prosodic dependency and grammatical dependency at the level of the phrase rather than the word' (p. 399) and 'may be treated as phrasal affixes' (p. 414). Minimally, Sm'algyax clitics share a range of typical clitic properties including the morphosyntactic properties of fixed ordering, syntactic dependency described by reference to syntactic elements (e.g. right edge of the modifier phrase, left of the phrasal head, or leftmost element of the predicate phrase), non-selectivity of the syntactic category of the host and syntactic scope over a phrase (following Aikhenvald 2003; Anderson 1992, 2005; Anderson and Zwicky 2003; Klavans 1985; Zwicky 1977, 1985; Zwicky and Pullum 1983).

Where the sets of Sm'algyax clitics differ markedly is in their degree of phonological dependency on an adjacent word. In this section we characterize the differences in dependency by considering stress, the phonological interaction of a clitic with its host, the ability to host other clitics, pausing and orthographic practices in determining word boundaries.

To begin, words typically have independent stress, whereas affixes and clitics do not. In Sm'algyax primary stress is a feature of lexical words only and it usually falls on the final syllable of the stem (Sasama 1995:51). None of the representative sets of clitics presented in the previous section carries primary stress and all are prosodically dependent on their host. However, a clause initial clitic may carry secondary stress, especially for emphasis or pragmatic reasons.

Comparing the phonological interaction of a clitic with its host to that of words and affixes provides another means for distinguishing the degree of phonological dependence of a clitic. In Sm'algyax phonological interaction between a stem and suffix includes stem-final lenition (as in (3) and (17)) and epenthesis (as in (7), (13) and (16)). We find that the predicative and attributive connectives, which are enclitics, trigger these same phonological processes. For example, stem-final lenition occurs with both the predicate connective *=a* and the attributive connective *=m* in (10) above, while in (14) vowel epenthesis occurs between the predicative connective *=sga* and its host *'wiisal<u>a</u>ks* 'make great fire':

[10] As Zwicky and Pullum (1983) identify that clitics, but not affixes, can be attached to material already containing clitics as one of the means for differentiating between clitics and affixes, the fact that a subjective dependent pronoun cliticizes to another clitic, albeit a rather non-prototypical clitic, is evidence that the subjective dependent pronouns are themselves clitics. It is relevant to note, however, that Klavans (1979: 71) has found that inflectional affixes can sometimes attach to clitics.

(14) *Ada wila 'wiisalaksüsga 'wii mediik.*
ada= wila= 'wii=sa-la̲ks[=sga 'wii= mediik.
and= then= great=make-fire[=CN.PRED great= grizzly.bear]
'And then great grizzly bear made a big fire.' (Mulder 1994:196)

Similarly, epenthetic vowels may occur between a subjective dependent pronoun and its preferred host as shown in (15) between the first part of the circumclitic *m*=...=*sm* '2PL.A' and its tense/aspect clause initial clitic host *dm* 'FUT':[11]

(15) *Ada dm düüsm ami midm sm gapt.*
ada= dm= düü-sm ami mi=dm =sm gap-t
and= FUT= die.PL-2PL.S if 2PL.A=FUT =2PL.A eat-3O
'And you will die if you eat it.' (Reeves 1992:12)

Turning to prefixes, there are a small number of derivational prefixes which undergo vowel harmony. While there is individual and dialectal variation, Sasama (1995) has found for Hartley Bay speakers that the derivational prefix *s*V- 'make, get' occurs as *sa*- before velar and glottal segments; *si*- before [j] and *sü*- elsewhere. In contrast, the possessive clitic, the modifier and locative clitics and the clause initial clitics do not undergo vowel harmony as is illustrated in (16), where the locative lexical proclitic *sa*= 'off' is followed by a vowel.

(16) *Siipga t'mbaayu ła sa oksu.*
siipk[=a t'mbaa-u] ła= sa= oks-u
hurt[=CN.PRED hip-1SG.POSS] PAST= off= fall.down-1SG.S
'My hip hurt when I fell down.' (Stebbins 2001a:164)

This along with the lack of epenthesis (as compared, for example, to the subjective dependent pronoun in (15) above) suggests that these three sets of clitics do not phonologically interact with their host, although they attach to it for their own prosodic realization.

Another marker of phonological independence is the ability to host other clitics. When the clause initial clitics host the subjective dependent pronouns they exhibit morphological complexity, which is a feature that is generally associated with words rather than affixes or clitics.

The occurrence of pausing after an element also points toward phonological independence, as pausing may occur between words, but typically not within an affix/clitic + stem combination. The clause initial clitics, but none of the other sets of clitics, have been observed as being a site for pausing (e.g. *ada…ada…* 'and…and…' from Stebbins 2003:401). This indicates a more word-like quality of this clitic set.

---

[11] Note here that =*sm*, the second part of the circumclitic *m*=...=*sm* '2PL.A', has been written disjunctively. This seems to have been the convention in the handful of relevant examples in this series of texts, but doesn't occur in any other works by Sm'algyax writers.

Finally, orthographic practice can give an indication as to how native speakers mentally divide up elements, although it certainly cannot be relied upon as the sole criterion for determining word boundaries. Typically, an affix/clitic + stem combination is written conjunctively (as a single word), whereas a word + word combination is written disjunctively (as separate words). With some types of clitics there is variation across native writers as to whether the clitic element and its prosodic host are written conjunctively or disjunctively. In contrast, other types of clitics are consistently written only conjunctively or only disjunctively. While Sm'algyax orthographic conventions are still being codified (cf. Stebbins 2001b), the observed practices of Sm'algyax writers with respect to different sets of clitics can be taken as further evidence of varying phonological dependence. What we find is that, consistently, predicative and attributive connectives are written conjunctively (as in (1)-(3), (5), (10), (12), (14), (16)-(18)), whereas clause initial clitics are written disjunctively (as in (1), (2), (4)-(5), (7)-(18)). The subjective dependent pronouns are written conjunctively when they are prosodically attached to a clause initial clitic (as in (5), (9)-(10), (13), (15), (17))[12] and disjunctively when they stand alone by being syllabic or by adding an epenthetic vowel to create a well-formed syllable and are prosodically dependent on the first lexical word of the clause (as in (11)-(12)). The possessive clitic and the modifier and locative clitics, however, tend to be written either conjunctively or disjunctively. For example, in (3) above the possessive clitic is written conjunctively while in (4) it is written disjunctively. Similarly, in (14) the modifier clitic *'wii* 'great' is written conjunctively when it is attached to the verb, but disjunctively when attached to the noun. Stebbins (1999:138) notes that "Beynon, probably the most prolific writer of Sm'algyax, in his manuscripts tends to attach the nearest lexical clitic to the stem and treat the others independently." What is important to observe here is the consistency across writers as to the conjunctive and/or disjunctive treatment of the various sets of clitics.

The observations about differences in phonological dependence can be summarized as shown in Table 2.

As we see here, Sm'algyax clitics appear to have three levels of phonological dependence. The predicative and attributive connectives are prototypical clitics and highly phonologically dependent as they are non-stress bearing, interact phonologically with their host, do not host other clitics, are not a site for pausing and are written conjunctively. The possessive clitic and the modifier and locative clitics are less prototypical and less phonologically dependent in that they do not interact phonologically with their host and are written either conjunctively or disjunctively. Clause initial clitics are the least prototypical and are the most phonologically independent, or word-like, as they can carry secondary stress, can host other clitics, can be a site for pausing, and are written disjunctively. The subjective dependent pronouns appear to pattern like the predicative and attributive connectives as prototypical clitics when they are hosted by a clause initial clitic, but like the clause initial clitics when they are prosodically dependent on the first lexical word of the clause.

[12] With the exception of the second part of the circumclitic *m*=...=*sm* '2PL.A' which has been written disjunctively as mentioned in the previous footnote.

TABLE 2. Summary of differences in phonological dependence

Sm'algyax Clitics	Location	Stress	Phonological Interaction with Host	Host other Clitics	Pausing	Orthographic Word
Predicative Connectives	enclitic	not stress bearing	epenthesis, stem-final lenition, coalescence	no	no	conjunctive
Attributive Connectives						
Possessive Clitic	proclitic	not stress bearing	none	no	no	conjunctive/ disjunctive
Modifier and Locative Clitics						
Clause Initial Clitics	proclitic	not primary stress bearing, but can carry secondary stress	none	yes	yes	disjunctive
Subjective Dependent Pronouns	clause initial clitic host: proclitic, enclitic, flexiclitic, circumclitic first lexical word host: proclitic	not stress bearing	epenthesis (with clause initial clitic host or to form a syllable)	no	no	clause initial clitic host: conjunctive first lexical word host: disjunctive

These three levels of phonological dependence can be taken as reflecting three different types of prosodic arrangements. By arrangement we refer to the manner in which a clitic is incorporated into the accentual structure of its host, since in order to be phonetically realized a word, affix or clitic needs to be part of an accentual unit. Turning to the theory of clitics, as Anderson (2005:46) remarks, "for any given host-plus-clitic combination, there are essentially four formal possibilities as to the prosodic structure that might result from combining them," whether this is at the level of the Phonological Word (PWd) or the Phonological Phrase (PPh).[13] These structures are shown in Figure 1:

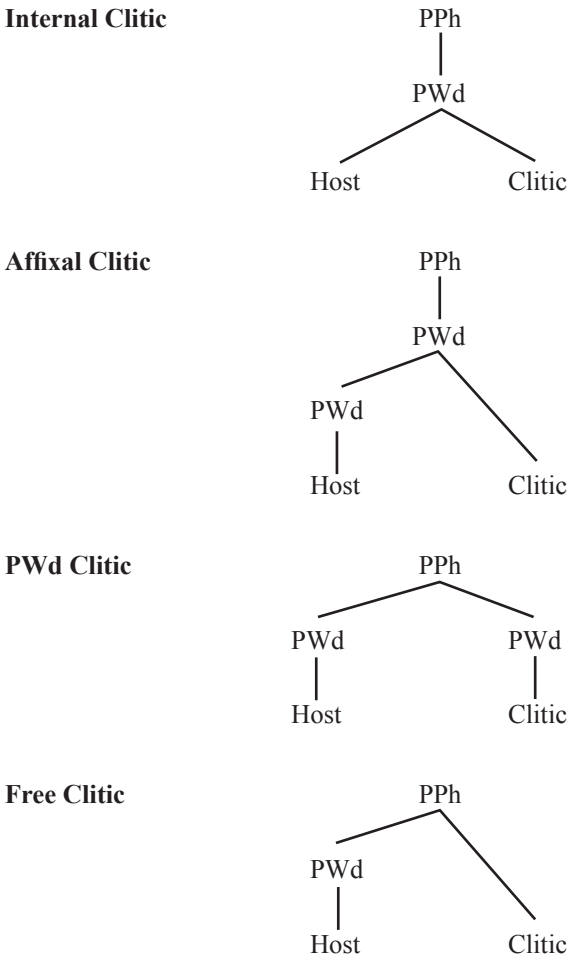


FIGURE 1. Prosodic structures of clitic attachment (Anderson 2005:46)

[13] Anderson attributes this observation originally to Selkirk (1995).

In the Internal Clitic prosodic structure the clitic becomes part of the same PWd as its host. The high level of phonological dependence of the predicative and attributive connectives on their host as evidenced by phenomena such as stem-final lenition, epenthesis and coalescence, suggests that both of these sets of clitics can be classified as Internal Clitics.

An Affixal Clitic prosodic structure is where the clitic joins its host, forming a recursive PWd structure.[14] The possessive clitic and the modifier and locative clitics are, arguably, Affixal Clitics. Since there is no phonological interaction with the host, such as vowel harmony, there is no phonological motivation for analyzing them as being incorporated into the same PWd as their host, and hence, unlike the predicative and attributive connectives, they cannot be analyzed as Internal Clitics. On the other hand, even though they are syllabic and can stand by themselves orthographically, they are not stress bearing and do not have the degree of phonological independence of the clause initial clitics. Thus, they do not constitute an independent PWd and instead are best analyzed as forming a PWd with their PWd host.

With a PWd Clitic prosodic structure, the clitic attaches to its host at the level of the PPh as an independent PWd. The clause initial clitics are examples of these since they are word-like in being able to host other clitics, carry secondary stress, are a site for pausing, and are written disjunctively. However, they are still clitics in that they have a fixed order and are phonologically dependent on the first lexical word in the clause, which is the head of the phonological phrase.

The subjective dependent pronouns, the remaining set of clitics, appear to be Internal Clitics when they are hosted by a clause initial clitic, but PWd Clitics when they occur independently and, along with any other clause initial clitics, form a prosodic unit with the first lexical word in the clause.

By taking into consideration the different prosodic structures of clitic attachment proposed in the theory of clitics, we are able to gain insight into the varied degrees of phonological dependency observed in the Sm'algyax clitics. At the same time, the Sm'algyax clitics provide evidence that Affixal, as well as Internal and PWd Clitics, actually occur in natural language and that properties such as stress placement, the phonological interaction of a clitic with its host, the hosting of other clitics, pausing and orthographic practices in determining word boundaries, can be used to distinguish among the various types of clitics.

2.3. POSITIONING CLITICS IN SM'ALGYAX. In classifying clitics in Sm'algyax we need to consider not only the phonological dimension, but also the morphosyntactic dimension of cliticization—that is, where the different sets of clitics appear. Following Klavans (1980, 1985) and Anderson (1992, 2005), we can characterize the positioning of Sm'algyax clitics in terms of their association with a syntactic domain, their location within this domain and their direction of attachment to their host.

What we have seen so far of the positioning of each of the representative sets of clitics can be summarized as shown in Table 3:

[14] This structure violates the non-recursivity criterion of the Strict Layer Hypothesis (Selkirk 1984), a point we return to in Section 3.4.

TABLE 3. Summary of differences in position

Sm'algyax Clitics	Syntactic Domain	Location within Domain	Direction of Attachment	Examples
Predicative Connectives	noun phrase	left edge	left (= enclitic)	(1), (2), (5), (10), (12), (14), (16)-(18)
Attributive Connectives	modifier phrase	right edge	left (= enclitic)	(2), (3), (10), (17)
Possessive Clitic	noun phrase	left edge	right (= proclitic)	(1), (3), (4), (5)
Modifier and Locative Clitics	noun phrase	left edge	right (= proclitic)	(2)-(6), (8)-(9), (14), (17)-(18)
	predicate phrase	left of head	right (= proclitic)	
Clause Initial Clitics	clause	left edge	right (= proclitic)	(1), (2), (4)-(5), (7)-(18)
Subjective Dependent Pronouns	predicate phrase	left-most element	• clause initial clitic host: right (=proclitic), left (=enclitic), left or right (=flexiclitic), both left and right (=circumclitic), • lexical word host: right (=proclitic)	(5), (9)-(10), (13), (15), (17) (11)-(12)

While the positioning of the predicative connectives, the attributive connectives, the possessive clitic and the clause initial clitics has been amply illustrated, that of the remaining two sets of clitics requires further discussion.

First, when modifying a verb, the modifier and locative clitics occur after any tense/aspect markers, discourse markers and lexical adverbs, and closest to the verb stem (as in (9) above). Consequently, their location within the predicate phrase can be described as to the left of the head. In contrast, when modifying a noun, the modifier and locative clitics occur to the left of the head and before any lexical modifiers that happen to be present (Stebbins 2003:410). As is illustrated in (17) (and in (2) and (3) above), the modifier *łgu=* 'little' attaches to another modifier within the NP to act phrasally on the noun head:

(17) *Niidzu* *wilt* *dibaada* *duusa*
niits-u [wil=t di=baa-t[=a duus][=a
see-1A [SUBORD=3A with=run-3o[=CN.PRED cat][=CN.PRED

łgudzagm *wüts'iin.*
łgu=[dzak=m] wüts'iin]]
little=[dead=CN.ATTR] mouse]]
'I saw a cat running with a dead mouse.' (Stebbins 2003:410)

Accordingly, the modifier and locative clitics are located with respect to the left edge of a noun phrase (following the predicative connective and the possessive marker if present).

In the case of the subjective dependent pronouns what we observed in Section 2.1 is that the direction of attachment is dependent on the availability of a tense/aspect marker as a preferred host and the person and number of the clitic. The first person singular and plural attach as proclitics to the tense/aspect marker whereas the second person plural attaches as a circumclitic. If there is no preferred host then they attach as proclitics to the first lexical word in the clause. The second person singular also attaches as a proclitic to the tense/aspect marker, but in its absence can either attach as an enclitic to another clause initial clitic (and hence is a flexiclitic) or as a proclitic to the first lexical word in the clause. The third person pronoun, on the other hand, attaches as an enclitic to a preferred tense/aspect marker host or another clause initial clitic if one is not present. While it has been observed as attaching as a proclitic to the first lexical word in the clause, this is rare.

In sum, Sm'algyax clitics are located with the syntactic domains of noun, modifier and predicate phrase, and clauses. Within these syntactic domains, the clitic may be located with respect to the left or right edge, the head or the leftmost element. The direction of attachment may be to the left, right, to the left or right, or both left and right.

2.4. THE DIMENSIONS OF SM'ALGYAX CLITICS. Drawing together the properties of the six sets of clitics that we have considered here, we can locate the dimensions of Sm'algyax clitics with respect to the broader typology of clitics. Along the phonological dimension, we observed that Sm'algyax clitics distinguish three different levels of phonological dependence in their attachment to a host which can be reflected in Internal Clitic, Affixal Clitic and PWd Clitic prosodic structure. In Sm'algyax, Internal Clitics are prototypical clitics, while Affixal Clitics are less so, and PWd Clitics are the least prototypical.

Along the morphosyntactic dimension, Sm'algyax clitics are located with respect to a variety of factors within a range of syntactic domains and exhibit a wide range of possibilities for the direction of attachment, as is summarized in Table 4. While there appears to be some preference for left edge proclitics within noun phrases and clauses, overall the syntactic domain, the location of the clitic within the domain and the direction of attachment, do not strongly correlate. In sharp contrast, with the exception of the subjective dependent pronouns, the direction of attachment does correlate with prosodic structure.

TABLE 4. Dimensions of Sm'algyax clitics

Sm'algyax Clitics	Prosodic Structure	Syntactic Domain	Location within Domain	Direction of Attachment
Predicative Connectives	Internal Clitic	noun phrase	left edge	left (= enclitic)
Attributive Connectives		modifier phrase	right edge	
Possessive Clitic	Affixal Clitic	noun phrase	left edge	right (= proclitic)
Modifier and Locative Clitics		noun phrase	left edge	
		predicate phrase	left of head	
Clause Initial Clitics	PWd Clitic	clause	left edge	right (= proclitic)
Subjective Dependent Pronouns	Internal Clitic (clause initial clitic host)	predicate phrase	left-most element	right (= proclitic), left (= enclitic), left or right (= flexiclitic), both left and right (= circumclitic)
	PWd Clitic (lexical word host)			right (=proclitic)

Internal Clitics may be located with respect to the left edge (e.g., the predicative connectives), the right edge (e.g., the attributive connectives) or the left-most element of the phrase (e.g., the subjective dependent pronouns with a clause initial clitic host), but, with the exception of the subjective dependent pronouns, the Internal Clitics attach as enclitics. In contrast, Affixal Clitics (e.g., the possessive clitic and the modifier and locative clitics) and PWd Clitics (e.g., the clause initial clitics and the subjective dependent pronouns with a lexical word host) are located either at the left edge, to the left of the head or with respect of the left-most element of the syntactic domain, but they all attach to the right as proclitics.[15]

[15] We return to the significance of this correlation in Section 3.2.

3. IMPLICATIONS FOR THE THEORY OF CLITICS.

3.1. A SKETCH OF ANDERSON'S OPTIMALITY THEORY APPROACH. In addressing the theory of clitics, Anderson (2005) argues that the phonological dimension of cliticization can best be captured in terms of independently motivated prosodic structures, couched within an Optimality Theory (OT) framework. Following traditional accounts, clitics are treated as prosodically deficient forms that in order to be pronounced must, like all phonetic content, be integrated into prosodic structure. This is formalized in the convention of Full Interpretation (Anderson 2005:39). How these prosodically deficient forms are incorporated into prosodic structure is captured through the principle of Stray Adjunction (Anderson 2005:13), while the precise prosodic structures that result from the integration of the deficient forms are independently motivated through the Prosodic Hierarchy of syllable, foot, phonological word, phonological phrase, etc. as developed in Selkirk (1984) and Nespor and Vogel (1986) and codified in the Strict Layer Hypothesis (Selkirk 1984, 1995). Where Anderson departs noticeably from traditional theoretical accounts of clitics is in treating the principles of the Strict Layer Hypothesis as violable constraints rather than inviolable rules on the basis of empirical evidence from a variety of languages.

In regard to the morphosyntactic dimension of cliticization, Anderson outlines a view of phrasal morphology within an OT framework that uses a limited set of constraint types, ranked relative to each other, to arrive at a characterization of the actual positioning of clitics within a language. While we do not give the details here, different constraint rankings can be specified to capture, for example, how some Sm'algyax clitics attach prosodically to the left but belong syntactically to the right while others attach prosodically within their syntactic domain.[16]

In the next two sections we look more closely at the principle of Stray Adjunction. In Section 3.2 we consider the evidence from Sm'algyax as to whether Stray Adjunction is language-specific or specific to particular clitics within a language, while in Section 3.3 we consider the interaction of Stray Adjunction and clause initial topics in Sm'algyax. Then in Section 3.4 we explore the further support that the Sm'algyax data provide for Anderson's treatment of the principles of the Strict Layer Hypothesis as violable constraints.

3.2. STRAY ADJUNCTION. In specifying how prosodically deficient forms such as clitics are incorporated into prosodic structure, the principle of Stray Adjunction identifies the direction of attachment of the stray material. An empirical question that arises in formulating this principle is whether the direction of attachment needs to be specified individually as a property of particular clitics, or whether it is a general property of a language. For Klavans (1985) directionality is clitic-specific, while for Anderson, directionality is "a matter that follows from the overall prosodic properties of the language, and which is not available for lexical specification with respect to individual items" (2005:60). He shows how the very few cases of purported clitic-specific attachment that have been identified in the literature

[16] See Sellers (2005) for a detailed elaboration of the relevant constraints and their rankings, including a proposal for an additional type of constraint to account for the subjective dependent pronouns attaching as a PWd Clitic.

can be analyzed as following from more general principles of the language's prosodic organization.

In Sm'algyax, as we saw in Section 2 and as is summarized in Table 5 below, the direction of attachment is specific to sets of clitics, with the exception of one clitic set, the subjective dependent pronouns, which has variable attachment specific to clitics within the set.

TABLE 5. Stray adjunction for Sm'algyax clitics

Sm'algyax Clitics	Prosodic Structure	Stray Adjunction
Predicative Connectives	Internal Clitic	left (= enclitic)
Attributive Connectives		
Possessive Clitic	Affixal Clitic	right (= proclitic)
Modifier and Locative Clitics		
Clause Initial Clitics	PWd Clitic	right (= proclitic)
Subjective Dependent Pronouns	Internal Clitic (clause initial clitic host)	right (= proclitic), left (= enclitic), left or right (= flexiclitic), both left and right (= circumclitic)
	PWd Clitic (lexical word host)	right (=proclitic)

We further observed that, with the exception of when the subjective dependent pronouns attach as Internal Clitics, the direction of attachment correlates with prosodic structure; Internal Clitics all attach to the left as enclitics, while Affixal and PWd Clitics all attach to the right as proclitics.

Though it is beyond the scope of the present chapter to formulate a complete analysis, it is reasonable to assume that for all but the subjective dependent pronoun Internal Clitics, the choice of leftward or rightward attachment can be analyzed as following from more general principles of Sm'algyax's prosodic organization. However, it is not at all clear how any such explanation could account for the behavior of the subjective dependent pronouns which attach as Internal Clitics since here the direction of attachment appears to be truly clitic-specific. As we outlined in Section 2.1, when there is a tense/aspect clitic present as a preferred host, the first person singular (*n*=), the first person plural (*dp*=) and second person singular (*m*=) attach rightward as proclitics, the second person plural (*m*=...=*sm*) attaches as a circumclitic and the third person (=*t*) attaches leftward as an enclitic.[17] Since these clitics are all introduced into the same syntactic position, it is not clear how any purely pro-

[17] Recall that if there is no preferred tense/aspect clitic host, then the pronouns can stand by themselves as PWd Clitics, while the second person singular and the third person can attach as Internal

sodic explanation nor how any explanation involving the interaction between prosodic and syntactic factors can account for the differences of directionality within this set of clitics. Instead, what it suggests is that this one set of clitics requires a clitic-specific indication of direction of attachment.

Thus, it seems that Sm'algyax provides empirical evidence in support of Klavans' (1985) view that the direction of attachment needs to be specified individually as a property of particular clitics, if only minimally for some sets of clitics in some languages.

3.3. CLAUSE INITIAL TOPICS . In K^{w}ak^{w}'ala and other northern Wakashan languages (two of which are neighbors of Sm'algyax) the principle of Stray Adjunction initially appears to influence the occurrence of clause initial noun phrase topics. Anderson (2005) finds that while these languages have a principle of Stray Adjunction to the left, languages without left edge clitic determiners allow a construction with a clause initial nominal representing a topic, whereas K^{w}ak^{w}'ala, which has left edge clitic determiners, does not. He concludes that: "the correlation between determiner systems and the possibility of clause initial topics make it clear that the absence of this latter construction in K^{w}ak^{w}'ala must be due to the impossibility of providing a host for the determiner clitics that appear at the left edges of nominals in this language" (Anderson 2005:22).

Interestingly, clause initial noun phrase topics occur reasonably often in Sm'algyax narratives, whose word order is otherwise very similar to that of K^{w}ak^{w}'ala. Like K^{w}ak^{w}'ala determiners the Sm'algyax predicative connectives, as we illustrated in Section 2.1, occur at the left edge of the noun phrase and attach as an enclitic to the preceding word.[18] However, when an S or A argument occurs as a clause initial topic, any predicative connective associated with it is no longer expressed, as is illustrated with an S topic in (7) and (8) above. When an O occurs as a topic, the predicative connective may either be unexpressed or occur at the right edge of the topic noun phrase (instead of in its usual position at the left edge of a noun phrase), as in (18):

(18)	*Ada*	*txa'nii*	*biyalsa*	*ksagawdit.*
	ada=	[txa'nii	biyals=a]	ksa=gawdi-t
	and	[all	star=CN.PRED]	out.of=be.all-3s

'And all the stars came out.' (Mulder 1994:204)

The implication for Anderson's analysis of K^{w}ak^{w}ala is that instead of taking a leftwards principle of Stray Attachment as preventing the occurrence of a clause initial topic, there may be other contributing factors to the existence or lack of such topics in a language.

---

enclitics to a neighbouring clause initial clitic. It is reasonable to assume that this behaviour can also be straightforwardly accounted for within a general prosodic account of Sm'algyax.

[18] The predicative connectives are not completely analogous with the K^{w}ak^{w}'ala determiners, however, as Sm'algyax also has a separate set of determiners which occur to the right of the nominal and are independent words. The Sm'algyax determiners have their own associated connective as well (always expressed as *=a*), which precedes the determiner but phonologically attaches as an enclitic to the preceding word.

3.4. PROSODIC STRUCTURE. The Prosodic Hierarchy provides a fundamental framework for accounting for the prosodic organization of language and as originally conceived by Selkirk (1984) the principles governing the use of the Prosodic Hierarchy within a language are codified as rules in the Strict Layer Hypothesis. Now, as we discussed in Section 2.2, the Prosodic Hierarchy predicts four prosodic structures of clitic attachment. However, of these four, two are ruled out by the Strict Layer Hypothesis; a Free Clitic involves a structure in which the clitic is dominated by a phonological phrase (PPh) without actually being part of a phonological word (PWd), while an Affixal Clitic involves recursion as a clitic joined to its PWd host to form a recursive PWd structure. Since these structures are motivated for a variety of languages, Anderson follows the idea that the principles of the Strict Layer Hypothesis are violable, and hold until there is a more pressing principle which overrides them in a language. It is this violable nature of the principles of the Strict Layer Hypothesis that is a major impetus for Anderson's formulation of a theory of clitics within an OT framework.

In Section 2 we have provided evidence that in Sm'algyax clitics either become part of the same PWd as their host (an Internal Clitic), have independent status as a PWd and form a phonological phrase with their host (a PWd Clitic), or join their host to form a recursive PWd structure (an Affixal Clitic). It is the existence of the last category, Affixal Clitics, in Sm'algyax which further calls into question the viability of conceiving of the Strict Layer Hypothesis as a set of inviolable rules and thus provides additional empirical support to Anderson's arguments for describing the prosodic structure of such clitics via a series of violable constraints within an OT framework.

4. CONCLUSIONS. In this chapter we have considered both the phonological and the morphosyntactic dimensions of cliticization as represented in six sets of clitics in Sm'algyax. In situating the complex behavior of Sm'algyax clitics within the broader typology of clitics, we have drawn on the theory of clitics.

Along the phonological dimension we have shown how the phonology of a language can be used to distinguish among the various types of clitics and by taking into consideration the different prosodic structures of clitic attachment proposed in the theory of clitics, we have been able to gain a clearer perception of the different degrees of phonological dependency observed in the data. At the same time, the existence of Affixal Clitics (alongside Internal and PWd Clitics) in Sm'algyax has implications for clitic theory; Affixal Clitics further call into question the viability of conceiving of the Strict Layer Hypothesis as a set of inviolable rules and provide additional empirical support to Anderson's (2005) arguments for describing the prosodic structure of clitics via a series of violable constraints.

Along the morphosyntactic dimension we have established that Sm'algyax clitics may be located with respect to the left or right edge, the leftmost element or the head within a range of phrasal and clausal syntactic domains. While in the syntactic domain, the location of the clitic within the domain and the direction of attachment do not strongly correlate, with the exception of one clitic set, the direction of attachment does correlate with prosodic structure; Internal Clitics all attach to the left as enclitics, while Affixal and PWd Clitics all attach to the right as proclitics. It is the subjective dependent pronouns, which have variable attachment specific to clitics within the set, that provide empirical evidence in support

of the view that the direction of attachment needs to be specified individually as a property of particular clitics (Klavans 1985), rather than following from more general principles of a language's prosodic structure (Anderson 2005).

In approaching theory from the field we find that not only does the fieldwork data have implications for linguistic theory but also that linguistic theory can provide insight to understanding the fieldwork data.

References

Aikhenvald, Alexandra Y. 2003. Typological parameters for the study of clitics, with special reference to Tariana. In R.M.W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word: A cross-linguistic typology*, 42-78. Cambridge: Cambridge University Press.

Anderson, Steven R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Steven R. 2005. *Aspects of the theory of clitics*. Oxford: Oxford University Press.

Anderson, Steven R. & Arnold M. Zwicky. 2003. Clitics. In William Frawley (ed.), *International encyclopedia of linguistics*, 2nd edn., vol. 1, 328-328. Oxford: Oxford University Press.

Boas, Franz. 1911. Tsimshian. In Franz Boas (ed.), *Handbook of American Indian languages,* 283-422. Washington, DC: Government Printing Office. Facsimile reproduction, 1981. Seattle: Shorey Book Store.

Boas, Franz. 1912. Tsimshian texts (New series). In Franz Boas (ed.), *Publications of the American Ethnological Society* 3, 65-285. Leyden: E.J. Brill.

Bolton, Priscilla, Marjorie Brown, Cheryl Keyworth & Sarah Shuster. 1984. *Goo magn sm nii maaxay*. Prince Rupert, BC: Sm'algyax Language Program, School District #52.

Dunn, John. 1978. *A practical dictionary of the Coast Tsimshian language* (National Museum of Man, Mercury Series, Canadian Ethnology Service Paper 42). Ottawa: National Museums of Canada.

Dunn, John. 1979a. *A reference grammar for the Coast Tsimshian language* (National Museum of Man, Mercury Series, Canadian Ethnology Service Paper 55). Ottawa: National Museums of Canada.

Dunn, John. 1979b. Pronominal concord in Coast Tsimshian. *International Journal of American Linguistics* 45. 224-231.

Dunn, John. 1995. *Sm'algyax: A reference dictionary and grammar for the Coast Tsimshian language*. Seattle: University of Washington Press.

Klavans, Judith. 1979. On clitics as words. *Chicago Linguistic Society* 15. 68-80.

Klavans, Judith. 1980. *Some problems in a theory of clitics*. Bloomington: Indiana University Linguistics Club.

Klavans, Judith. 1985. The independence of syntax and phonology in cliticization. *Language* 61. 95-120.

Mulder, Jean G. 1994. *Ergativity in Coast Tsimshian (Sm'algyax)*. Berkeley: University of California Press.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Reeves, Henry. 1992. *Adawga gant wilaaytga gyetga suwildook (Rituals of respect and the sea otter hunt)* (Suwillaay'msga Na Ganiiyatgm, Teachings of our Grandfathers Series, vol. 2). Prince Rupert, British Columbia: Prince Rupert School District #52.

Sasama, Fumiko. 1995. *Coast Tsimshian plural formation with phonological introduction*. Sapporo, Japan: Hokkaido University MA thesis.

Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, Elisabeth O. 1995. The prosodic structure of function words. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in optimality theory* (University of Massachusetts Occasional Papers in Linguistics 18), 439-469. Amherst: Department of Linguistics, University of Massachusetts.

Sellers, Holly. 2005. *Cliticisation in Sm'algyax (Coast Tsimshian): A phonological and morphosyntactic constraint-based account*. Melbourne, Australia: University of Melbourne honors thesis.

Stebbins, Tanya. 1999. *Issues in Sm'algyax (Coast Tsimshian) lexicography*. Melbourne, Australia: University of Melbourne dissertation.

Stebbins, Tanya (compiler). 2001a. *Sm'algyax learner's dictionary,* vols. 1-5. Prince Rupert: Ts'msyeen Sm'algyax Authority.

Stebbins, Tanya. 2001b. Emergent spelling patterns in Sm'algyax. *Written Language and Literacy* 4. 163-194.

Stebbins, Tanya. 2003. On the status of intermediate form classes: Words, clitics, and affixes in Sm'algyax (Coast Tsimshian). *Linguistic Typology* 7. 383-415.

Zwicky, Arnold M. 1977. *On clitics*. Bloomington: Indiana University Linguistics Club.

Zwicky, Arnold M. 1985. Clitics and particles. *Language* 61. 283-305.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59. 502-513.

Jean Mulder
jmulder@unimelb.edu.au

Holly Sellers
hollyas@gmail.com

4

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp.57-89
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4451

Noun class and number in Kiowa-Tanoan: Comparative-historical research and respecting speakers' rights in fieldwork[1]

Logan Sutton

University of New Mexico

The Kiowa-Tanoan family is known to linguists by two characteristic features: a) a package of complex morphosyntactic structures that includes a typologically marked noun class and number marking system and b) the paucity of information available on the Tanoan languages due to cultural ideologies of secrecy. This paper explores both of these issues. It attempts to reconstruct the historical noun class-number system based on the diverging, yet obviously related, morphosemantic patterns found in each of the modern languages, a study that would be greatly benefited by fieldwork and the input of native speakers. At the same time, it reviews the language situation among the Kiowa-Tanoan-speaking communities and what some of the difficulties are in doing this kind of fieldwork in the Pueblo Southwest, touching on the myriad complex issues involving the control of information and the speech communities' rights over their own languages as well as the outside linguist's role in such a situation. The paper underscores these points by using only language data examples from previous field research that are already available to the public so as not to compromise native speakers' sensitivity to new research on their languages.

1. INTRODUCTION. The Kiowa-Tanoan languages exhibit a typologically unusual system for dividing nouns into classes and marking number. Generalizing over all of the languages, nouns are distributed across four noun classes which have largely been determined by the grammatical number denoted by the basic form as opposed to a form with an inverse suffix which 'reverses' the basic number, e.g. basic singular becomes inverse plural, or basic plural becomes inverse singular. Although the five languages (Kiowa, Tewa, Northern Tiwa, Southern Tiwa, and Towa) differ to a greater or lesser extent among their systems, they demonstrate numerous similarities that are suggestive of the organization of the his-

[1] The author thanks the following colleagues at the University of New Mexico for helpful comments on various versions of this paper and its presentation for the 2007 Workshop for American Indian Languages at University of California, Santa Barbara: Evan Ashworth, Melissa Axelrod, Brittany Kubacki, Bill Croft, Steven Menefee, and Caroline Smith. Grateful acknowledgments also go to the editors of the present volume—Andrea Berez, Daisy Rosenblum, and Jean Mulder—and three anonymous reviewers for their valuable input, comments, and encouragement on earlier drafts. The author bears responsibility for any mistake still lingering in these pages. Last but far from least the author wishes to thank those native speakers of Tanoan languages that he has had the pleasure to work with and those who have shared their beautiful languages with previous researchers, *sine qua non*.

torical system from which the synchronic ones derive. This paper represents a preliminary attempt to reconstruct the semantic structure of that proto-system. The following discussion will outline the synchronic patterns found in each of the modern languages, compare and contrast the major differences between the systems in order to isolate the likely historical conservations from the innovations, and finally hypothesize what some of the semantic bases of that system may have been.

1.1. DATA. The published literature on the Kiowa-Tanoan languages is rather sparse, most of it having been produced in scattered articles before 1970. Several unpublished dissertations based upon fieldwork—from the Tragers' work on Northern Tiwa in the 1930s-60s to Yumitani's work on Towa in the 1980s-90s—have filled in many gaps in our knowledge about the Kiowa-Tanoan languages. One main reason for this lack of published comprehensive descriptions has been a desire by the speakers of the languages—especially among the Pueblo communities—not to have their languages written down and their privacy invaded by outsiders. As such, the author of this paper limits himself to the available data in print, although some of the analyses and conclusions have been informally enriched and verified from his own experience in working with native speakers.

The present collection extols the many benefits of deriving linguistic analyses from data collected in the field. It may therefore be surprising that I base the hypothesis and the conclusions of this paper on data collected by somebody else. Despite the fair amount of information that can be gleaned from the work that has been done to date on Kiowa-Tanoan languages, a great deal more fieldwork is still required before a comprehensive treatment of the languages can be produced. This is especially true for an analysis of the noun class systems discussed here, which ideally requires the collection of huge quantities of vocabulary with concomitant grammatical features. This can only come from eliciting from native speakers word lists and narratives to see the usage of these words. However, with all of the virtues of fieldwork—indeed, most linguistic work is ultimately owing to native speakers sharing their languages—there are restrictions that must be recognized and respected. These restrictions do not derive from the quality of the data, but rather from the respective roles of the linguist and the native speakers in the fieldwork situation. It is easy for the naïve linguist to think about going to the field, finding consultants, doing elicitation, and coming back with data on which to base his or her studies, but what about the native speaker? Linguistic fieldwork by its nature requires working with other human beings. What the linguist needs to take into consideration is the fact that these other human beings also have a say and may have their own goals in the fieldwork process. This can realize itself in at least two ways: a) what the speakers are willing to divulge, and b) what the speakers think the linguist can do for them.

The Pueblos are renowned among linguists and anthropologists for their tradition of secrecy regarding aspects of their culture. With respect to language, this has meant a reluctance or downright refusal to allow the languages to be transcribed to the written medium and to be learned by outsiders. For those linguists who have had the good fortune to work with Pueblo speakers, this perspective has meant that any data collected in the community should not be easily accessible to other outsiders. This is obviously going to put severe restrictions on publication, the product towards which most academics are striving and,

indeed, are often obligated to produce in order to receive funding. Those works that have been published on Pueblo languages have often been done so without the communities' consent or with only the consent of a handful of speakers who prefer to remain anonymous (the means by which George Trager produced his numerous articles on Taos Tiwa). Of course once the data are collected, the linguist could choose to go against the wishes of the community and publish anyway, but the unscrupulousness of this act goes without saying.

The other question that is raised concerns the benefit the linguistic work has to the speech community. It is standard practice to pay the consultants with whom one works, but it must be borne in mind that this payment is a recompense for the speaker divulging information concerning part of his or her cultural heritage. In the case of most Native American communities—as well as other communities all around the globe—this heritage is in danger of disappearing under the weight of external social pressures. Members of the communities often recognize this fact and want to take action to prevent this loss. In this the linguist can help more than by simply providing a little income. Whatever the linguist's own personal linguistic goals in doing the fieldwork, s/he can—and should—assist the members of the community in their own linguistic endeavors insofar as s/he is able. This often means producing accessible pedagogical materials for the language and teaching native speakers how to analyze their own language in order to facilitate an intra-community language revitalization program. It is even the case nowadays that language revitalization efforts are being initiated within the speech community, with linguists being invited to assist in this effort. Such collaborative linguistic efforts can be of benefit to all parties; but, however the fieldwork is instigated, the linguist needs to keep in mind just what kind of payment s/he actually owes the speech community.

These issues lie at the heart of working with the Kiowa-Tanoan family of languages. Out of respect for the desire of Pueblo communities to have discretion over their own languages, the author has not drawn any data from his own fieldwork, grounded as it has been in collaborative community-based language revitalization projects. All data for this study therefore come from available articles, dissertations, and notes that are publicly accessible. This decision is not without its own moral dilemmas in respecting the rights of speakers over their own languages, but the author will leave it to the reader to decide whether it is justifiable or not.

2. NOUN CLASS AND NUMBER IN THE MODERN KIOWA-TANOAN LANGUAGES. This section will outline the synchronic formal and semantic features of the noun class systems of the five Kiowa-Tanoan languages. The languages are presented in the following order: Kiowa (section 2.1), Towa (section 2.2), Tewa (section 2.3), and Northern and Southern Tiwa (section 2.4).

2.1. KIOWA. Compared to the other Tanoan languages, Kiowa (Anadarko, Carnegie, and Lawton, Oklahoma) has had a huge amount of fieldwork done on it and has received some description in a sizable number of publications, from the lengthy word list in Mooney (1898) to Harrington's impressive (1928) vocabulary to Watkins' excellent (1984) grammar, among others. Unlike the Pueblo groups, the Kiowa belong to the very different Plains

cultural area, which does not have—as a general rule—as stringent a code of secrecy regarding language.

The Kiowa system maintains a basic-inverse number marking strategy delineating four classes into which nouns fall with respect to this strategy. The intricacies of the noun class system have been continuously analyzed and reanalyzed by scholars (Harrington 1928, Wonderly et al. 1954, Merrifield 1959, Trager 1960, Watkins 1984), rendering it the best described of all the Kiowa-Tanoan noun class systems. As will be seen in the next two sections, this system is similar to those in Towa and Tewa, but differs significantly in certain core aspects.

Noun class and number are formally marked in two main areas of the grammar: in the pronominal prefixes of the verb—portmanteau morphemes that index the person and number of the core arguments of the verb—and in the number suffix on nouns. Other elements in a noun phrase, notably demonstratives and some adjectives, are also marked for number. Because it most clearly highlights the basic versus inverse number contrast that forms the basis of the noun class distinctions, the number suffix on nouns will be illustrated first.

The inverse suffix in Kiowa has many allomorphs, the most productive of which—and therefore taken to be the most representative—is {*-gɔ́*}. The other forms (*-mɔ́*, *-bɔ́*, *-dɔ́*, *-tɔ́*, *-gú, -óy*, *-óp*, falling tone) are determined both phonologically and lexically, but do not appear to be conditioned by noun class in any way (Watkins 1984: 80). The number semantics of this morpheme depends upon the class of the noun to which it attaches. When attached to a class I noun, it indicates plural number, while its absence (the basic form) indicates either singular or dual. On a class II noun, it indicates singular, while the basic stem is both dual and plural. With class III nouns, the inverse marks singular and plural, leaving the basic to indicate dual number. Class IV is *prima facie* based on the absence of inverse marking: the basic form is used no matter the grammatical number. (1) illustrates this number marking, the distinctions being highlighted (data from Watkins 1984):

(1)		Basic	Inverse[2]
	I	*tógúl* 'young man' (**SG/DU**)	*tógú:**dɔ́*** 'young men' (PL)
	II	*ɔ̀nsó:* 'feet' (**DU/PL**)	*ɔ̀ns**ôy*** 'foot' (SG)
	III	*álɔ̀:* 'apples' (**DU**)	*álɔ̀:**bɔ̀*** '*apple*, apples' (SG/PL)
	IV	*hóldà* 'dress, dresses' (**SG/DU/PL**)	

As first pointed out in Merrifield (1959) and later reconfirmed in Watkins (1984), noun class membership cannot be based solely upon the distribution and semantics of the inverse suffix. The pronominal indexation prefix on the verb appears to play a much more sig-

[2] Abbreviations used in this paper are: BAS='basic number'; DU='dual'; IMP='imperative'; INDF='indefinite'; INV='inverse number'; PL='plural'; PST='past'; SG='singular'. For the pronominal prefixes 1, 2, 3 are for person; S, D, P are for number; A, B, C in Tiwa are for noun class marking; INAN in Tewa is for inanimate singular/plural subject; the order in di/tri-valent pronominal prefixes is Subject:Dative:Object.

nificant role in concisely determining the class (and subclass) of nouns.[3] For third person arguments, there are four formal number distinctions in the pronominal prefix paradigm:[4] singular (*Ø-*), dual (*ę̀-*), plural (*gyà-*), and inverse (*è-*). It is by the distribution of these number distinctions in correlation with the inverse suffixation pattern on nouns by which the noun classes may be most succinctly analyzed.[5] The following examples illustrate the number distinctions for each of the noun classes with respect to the pronominal indexation.

The indexation for class I nouns matches the pattern shown with the inverse suffix: singular takes the singular prefix, dual takes the dual prefix, and plural takes the inverse prefix. Note that the demonstratives are also marked for basic versus inverse number.

(2) (Watkins 1984:97)

a. *ę́:dè* *sân* *Ø-khóp-dɔ́:*
this.**BAS** child.**BAS** **3S-**hurt-be
'This child is sick.'

b. *ę́:**dè*** *sân* ***ę̀**-khóp-dɔ́:*
this.**BAS** child.**BAS** **3DU-**sick-be
'These children (DU) are sick.'

c. *ę́:**gɔ̀*** *są̂:dɔ* ***è**-khóp-dɔ́:*
this.**INV** child.**INV** **3INV-**hurt-be
'These children (PL) are sick.'

Class III similarly matches the inverse marking pattern on nouns: both singular and plural take the inverse prefix while dual takes the dual prefix. Compare (3a) and (3c), both of which take an inverse suffix on the noun and an inverse pronominal prefix, being differentiated only by the number suppletive verb stem.[6] (3b), being dual, takes a dual pronominal prefix and the basic form of the noun.

---

[3] Aside from deriving subclasses of classes II and IV, this also has the effect of reclassifying a few nouns that do not take an overt inverse suffix, e.g. *t'áp* 'deer'. These nouns are then solely classed based upon the pattern of pronominal indexation.

[4] Because of the complexity of the pronominal prefix system, which encodes subject, direct object, and, to a degree, indirect object, only the intransitive system—which only marks subject—will be discussed here.

[5] It needs to be noted here that the term 'inverse' is being used in two distinct, but related, ways. One usage pertains to the paradigmatic contrast between 'inverse' and 'non-inverse' (i.e. basic number) that serves to distinguish the noun classes. In other words it refers to a pattern that has formal realizations somewhere in the grammar, most notably in the number suffix on nouns which only makes this two way distinction. The other usage pertains to the formal markers of inverse number themselves, whether the suffix on the nouns and modifiers or the pronominal prefix.

[6] Several verbs in Kiowa-Tanoan languages have suppletive singular versus non-singular stem pairs, as seen in (3-7) and many others throughout this paper. Dual number takes the singular stem

(3) (adapted from Watkins 1984:89)

a. *ę́:gɔ̀ ɔ́:dɔ́ è-k'ɔ́:*
there hair.**INV** **3INV-**be.lying.SG/DU
'There's a strand of hair lying there.'

b. *ɔ́l ę́:gɔ̀ **ę̀-**k'ɔ́:*
hair.**BAS** there **3D-**be.lying.SG/DU
'There are two strands of hair lying there.'

c. *ɔ́:**dɔ́** hóldàp **è-**k'úl*
hair.**INV** dress.on **3INV-**be.lying.PL
'There's some hair on the dress.'

Classes II and IV are different insofar as they are divided into subclasses based upon how these pronominal prefixes are applied. Class II has two subclasses. In both, singular is marked with the inverse prefix and dual is marked with the dual prefix, as expected from the suffix on the noun. The difference lies in the plural. Class IIa takes the plural prefix for indicating plural, whereas class IIb takes the singular prefix. The sentences in (4) illustrate the use of the noun stem *á*: which, as a IIa noun, means 'pole' and, as a IIb noun, means 'tree'.

(4) (Watkins 1984:86-87)

a. (IIa / IIb)
*á:-**dɔ́** **è-**cél*
stick-**INV** **3INV-**stand.SG/DU
'A pole/tree is standing there.'

b. (IIa / IIb)
*á: **ę̀-**cél*
stick **3DU-**stand.SG/DU
'Poles/Trees (2) are standing there.'

c. (IIa)
*á: **gyà-**sɔ́l*
stick **3PL-**stand.PL
'Poles (3+) are set up there.'

---

with some verbs—as with *cél* 'stand' here—and the plural stem with others. Transitive verbs with stem suppletion reflect the number of the object, not the subject.

d. (IIb)
á: **Ø-***sɔ́l*
stick **3S-**stand.PL
'Trees (3+) are standing there.'

Class IV divides into three subclasses based on the pronominal prefixes. But, just as the inverse suffix never occurs on any class IV noun stems, so too does the inverse pronominal prefix never index a class IV argument; only the singular, dual, and plural prefixes are used. Classes IVa and IVb both use the singular prefix for singular and the dual for dual. As with the subclasses of II, the difference is in the marking of the plural. Class IVa marks the plural with the plural prefix and IVb marks the plural with the singular prefix. Class IVc differs from the other two in that only the plural prefix is used, whatever the semantic number is. (5) demonstrates IVa.[7] (6) and (7) show IVb and IVc respectively, using the noun *tó:* which, as a IVb noun, means 'house' and, as a IVc noun, means 'tepee'.

(5) (adapted from Watkins 1984:90)
a. (IVa)
ɐ́:dè *c'ó:* *ɐ́:gɔ̀* **Ø-***cél*
this.**BAS** rock there **3S-**be.sitting.SG/DU
'This rock is sitting there.'

b. (IVa)
*ɐ́:**dè*** *c'ó:* *ɐ́:gɔ̀* ***è-****cél*
this.**BAS** rock there **3D-**be.sitting.SG/DU
'These rocks (2) are sitting there.'

c. (IVa)
*ɐ́:**dè*** *c'ó:* *ɐ́:gɔ̀* ***gyà-****sɔ́l*
this.**BAS** rock there **3P-**be.sitting.PL
'These rocks (3+) are sitting there.'

(6) (Watkins 1984:90-91)
a. (IVb)
tó: **Ø-***cél*
house **3S-**stand.SG/DU
'There is a house standing there.'

[7] These sentences in (5) are constructed by the author. The only examples given in the literature of this paradigm occur with transitive pronominal prefixes (see Watkins 1984:90), but in the interest of simplicity and clarity, forms with intransitive prefixes are given here.

b. (IVb)
tó: ***ẹ̀-****cél*
house **3D-**stand.SG/DU
'There are houses (DU) standing there.'

c. (IVb)
tó: ***Ø-****sɔ́l*
house **3S-**stand.PL
'There are houses (PL) standing there.'

(7) (Watkins 1984:90-91)
(IVc)
tó: ***gyà-****sɔ́l*
house **3P-**stand.PL
'There is a tepee standing there/There are tepees (DU/PL) standing there.'

Table 1, adapted from Watkins (1984:79), summarizes the formal marking of noun classes in Kiowa.

TABLE 1. Kiowa Noun Classes

	Singular		Dual		Plural	
	Inv sfx	Index	Inv sfx	Index	Inv sfx	Index
Class I		*Ø-*		*ẹ̀-*	*-gɔ́*	*è-*
Class IIa	*-gɔ́*	*è-*		*ẹ̀-*		*gyà-*
IIb	*-gɔ́*	*è-*		*ẹ̀-*		*Ø-*
Class III	*-gɔ́*	*è-*		*ẹ̀-*	*-gɔ́*	*è-*
Class IVa		*è-*		*ẹ̀-*		*gyà-*
IVb		*Ø-*		*ẹ̀-*		*Ø-*
IVc		*gyà-*		*gyà-*		*gyà-*

To clarify the system in short, the language has a three-way number distinction between singular, dual, and plural. Crosscutting this system is a paradigmatic contrast between 'basic' and 'inverse' numbers, where the specific number semantics of 'basic' and 'inverse' are determined by the noun class to which the relevant noun belongs (e.g. 'inverse' is plural in class I, singular in class II, and singular or plural in class III). These distinctions are realized via a four-way morphological distinction in the pronominal prefixes

between singular, dual, plural, and inverse. The inverse morphological marking is used to realize formally the number that corresponds to the paradigmatic usage of the 'inverse' vs. 'basic' pattern, i.e. the inverse prefix *è-* will mark plural number with a class I noun, singular with a class II noun, and both singular and plural with a class III noun. The other morphological markers—singular, dual, and plural—then serve to realize the singular-dual-plural number distinction in the 'non-inverse' part of the paradigm. For example, since both the singular and dual of class I nouns are paradigmatically 'basic', the language is still able to disambiguate singular versus dual number formally. This complex system—and unfortunately confusing terminology that has accompanied it in the literature—is made no simpler by the fact that there are other semantic categorization patterns also feeding into the noun class distinctions.

Although the semantic basis for the distribution of nouns among the four classes is not entirely predictable, there are some very prevalent patterns. The most obvious feature is that all nouns with animate referents fall under class I. This includes both humans and higher and lower animals, and applies to terms for individuals, occupations, and groups (e.g. tribes). There do not appear to be any exceptions to this rule. Inanimates, on the other hand, are spread out across all four classes with no immediately transparent pattern to their distribution. This distribution is not equal, however. There are relatively few inanimate nouns in class I, these mostly being body parts (8a), objects capable of independent motion (8b), certain prominent objects in nature (8c), and certain objects made by humans that could be construed as instruments (8d), although further analysis is needed for verification of this (data from Watkins 1984).

(8) a. *t'ɔ́:de* 'ear'
tá:de 'eye'
tʰén 'heart'
mɔ̀nc'ó 'fingernail'

b. *khɔ̂:* 'car'
k'ɔ́dál 'vehicle'

c. *páy* 'sun'
p'ɔ́: 'moon'

d. *k'ɔ̀:* 'knife'
t'ɔ́: 'spoon'

Class III is in fact a closed class with no more than four to eight members, all of which are inanimate. Watkins (1984) gives the words in (9) as the comprehensive set.

(9) *álɔ̃:* 'apple; plum'
thǫ́t'ólǫ̃: 'orange'
k'ɔ̃n 'tomato'
ɔ́l 'hair'

Three out of four of these are roundish fruits[8] which share no apparent semantic relation with the fourth member 'hair'.[9]

Classes II and IV each also show patterns to their membership, with the former including "inanimate but tangible objects" and the latter "abstract inanimates" (Watkins 1984:92).[10] The further distinction into subclasses in II and IV appears to be largely based on the actual semantic construal of number. Watkins (1984) illustrates the difference between IIa and IIb—which only differ in the plural—as being a contrast between distributive and collective plural number. With IIa nouns, plural indexation is used to convey a bounded heterogeneous construal of number, a number of discernibly discrete objects. With IIb nouns, though, singular indexation is used to convey a bounded homogeneous construal, a collective of objects acting as a single unit. This is exemplified by (4), repeated here. The stem *á:*, in class IIa means 'pole, stick', but in IIb denotes 'tree' (i.e. a collective of sticks).

(4) (Watkins 1984:86-87)

a. (IIa / IIb)
*á:-**dɔ*** *__è-__cél*
stick-**INV** **3INV-**stand.SG/DU
'A pole/tree is standing there.'

b. (IIa / IIb)
á: *__ẹ̀-__cél*
stick **3DU-**stand.SG/DU
'Poles/Trees (2) are standing there.'

c. (IIa)
á: ***gyà-****sɔ́l*
stick **3PL-**stand.PL
'Poles (3+) are set up there.'

d. (IIb)
á: **Ø-***sɔ́l*
stick **3S-**stand.PL
'Trees (3+) are standing there.'

---

[8] Wonderly et al. 1954 lists a few more words, also denoting fruits (*sané'e* 'blackberry', *t'áp'alɔgɔ* 'wild cherry', *álɔ:gu'k'ó:* 'lemon; orange', *alɔ:sɔhyé* 'plum').

[9] Takahashi (1984) tentatively suggests that 'hair' may be the prototype for this class—on the basis of the Kiowa custom of wearing two braids in the hair—with the semantic extension to fruits representing 'things growing out of a main body' (p. 31, 37). Thanks to Hiroto Uchihara for bringing this to my attention.

[10] Class IV may also include most—if not all—non-count nouns, but no source makes this explicit.

The same contrast obtains between the subclasses of IV, IVb being a collective, and IVc a distributive, plural (Watkins 1984:90-91). (6)-(7), repeated here, illustrate this difference. The form *tó:* means 'house' when a IVb noun, a set of parts viewed as a collective unit, but 'tepee' when IVc, a single unit viewed as distributed multiple parts.

(6) (Watkins 1984:90-91)

a. (IVb)
tó: **Ø-***cél*
house **3s-**stand.SG/DU
'There is a house standing there.'

b. (IVb)
tó: ***ę̀-****cél*
house **3D-**stand.SG/DU
'There are houses (DU) standing there.'

c. (IVb)
tó: **Ø-***sɔ́l*
house **3s-**stand.PL
'There are houses (PL) standing there.'

(7) (Watkins 1984: 90-91)
(IVc)
tó: ***gyà-****sɔ́l*
house **3P-**stand.PL
'There is a tepee standing there/There are tepees (DU/PL) standing there.'

Class IVc always marks its members as plural, nouns such as *kut* 'book', *tó:* 'tepee', and *hóldà* 'dress, shirt' being construed as composed of multiple constituent parts (Merrifield 1959: 270, Watkins 1984:91) no matter how many of the overall units there are.

Despite some early controversy over the relatedness of Kiowa to the other Tanoan languages (cf. Trager & Trager 1959, Whorf & Trager 1937), a glance at the noun class system of Kiowa described here and the following descriptions of the Tanoan languages should leave the reader with no reservations over the connection, even without detailed sound correspondences and formal reconstructions (cf. Hale 1962, 1967; Watkins 1977, 1978, 1982, 1996). Towa, being the most similar to Kiowa in its noun class system, will be described next.

2.2. TOWA. Towa (Pueblo of Jemez, New Mexico), like Kiowa, uses a basic-inverse number marking system, the patterns of which analysts have used to divide all nouns into four classes comparable to those in Kiowa. The primary sources for these data are two dissertations (Sprott 1992, Yumitani 1998), which explicitly state the noun classes for several, although still a very limited set of, nouns. Until these two works, Towa remained the most

poorly described of the Pueblo languages. Jemez is one of the Pueblos most closed to linguistic work, but fortunately has a very high rate of transmission to the younger generations. This renders the need for explicit language revitalization efforts and studies by outside researchers a low priority from the official community's perspective.

The Towa inverse suffix is *-š*,[11] and like the inverse in Kiowa, its semantics is dependent upon the noun to which it attaches. Class I nouns use the basic form for singular and the inverse for dual and plural. Class II has the inverse in the singular and dual and the basic form in the plural. Class III is basic for singular and plural and inverse for dual. Class IV nouns never take the inverse suffix. (10) lays out the Towa pattern (data from Yumitani 1998).

(10)		Basic	Inverse
	I	*hí:* 'person' (SG)	*hí:míš* 'people' (DU/PL)
	II	*pǫ́:* 'roads' (PL)	*pǫ́:š* road; roads (DU)
	III	*kʷǫ́:* 'tooth' (SG); 'teeth' (PL)	*kʷǫ́:š* 'teeth' (DU)
	IV	*p'ê* 'water'	

Notice that the dual in Towa is inverse, the opposite of what is found in Kiowa, which shows the basic number in the dual. Concomitant to this difference, class III is basic in the singular and plural compared to inverse in Kiowa. This comparison will be further discussed in section 3.1.

Although the distribution of the inverse suffix is more predictive for noun class in Towa than it is in Kiowa, the pronominal prefix system again appears to be the best basis for assigning nouns to classes. To illustrate using the intransitive third person prefixes, there are again four formal numbers distinguished: singular (*Ø-*), dual (*i̧l-*), plural (*il-*), and (non-dual) inverse (*e-*). Although the dual always takes the inverse suffix on nouns, it takes the dual prefix in its pronominal indexation. The inverse prefix listed here occurs when the argument marked as inverse is non-dual. The plural prefix is primarily used only with some class II and III nouns, where it contrasts with the use of the singular prefix when there are plural referents. The following sentences demonstrate the noun classes by use of the inverse suffix and pronominal prefixes (examples from Yumitani 1998:100).

Class I: Basic in the singular with a singular pronominal prefix.
Inverse suffix on the noun with a dual pronominal prefix in the dual.
Inverse suffix and pronominal prefix in the plural.

(11) a. *pí̧ę:* **Ø-***ší*
deer **3s-**fall.SG/DU
'A deer fell off.'

[11] The language maintains a CV(:) syllable structure and the *-š* is frequently omitted when not utterance-final. Since the suffix does cause phonological changes in the preceding and following syllables, however, this presence of the inverse marker is still preserved (cf. Yumitani 1998:68-81).

b. *pą́ę́:-**š*** ***į̨il-**ší*
deer-**INV** **3D-**fall.SG/DU
'Deer (2) fell off.'

c. *pą́ę́:-**š*** ***e-**tʸí*
deer-**INV** **3INV-**fall.PL
'Deer (3+) fell off.'

Class II: Inverse suffix and pronominal prefix in the singular.
Inverse suffix and dual pronominal prefix in the dual.
Basic in the plural with a singular pronominal prefix (but see below for use of the plural prefix).

(12) a. *pǫ́:-**š*** ***e-**ší*
drum-**INV** **3INV-**fall.SG/DU
'A drum fell off.'

b. *pǫ́:-**š*** ***į̨il-**ší*
drum-**INV** **3D-**fall.SG/DU
'Drums (2) fell off.'

c. *pǫ́:* ***Ø-**tʸí*
drum **3S-**fall.PL
'Drums (3+) fell off.'

Class III: Basic in the singular with a singular pronominal prefix.
Inverse suffix in the dual with a dual pronominal prefix.
Basic in the plural with a singular pronominal prefix (but see below for use of the plural prefix).

(13) a. *bélá* ***Ø-**ší*
bread **3S-**fall.SG/DU
'(A loaf of) bread fell off.'

b. *bélé-**š*** ***į̨il-**ší*
bread-**INV** **3D-**fall.SG/DU
'(Loaves of) bread (2) fell off.'

c. *bélá* ***Ø-**tʸí*
bread **3S-**fall.PL
'(Loaves of) bread (3+) fell off.'

Class IV: No inverse marking; singular pronominal prefix.

(14) *ʼǫ́:ʼe* **Ø-***tʸí*
sugar **3S-**fall.PL
'Sugar fell off.'

The difference between the singular and the plural pronominal indexation in the plurals of classes II and III appears to be along similar lines as the difference between classes IIa and b and between IVa and b in Kiowa. However, according to Yumitani (1998: 105-6), it is the plural prefix that denotes a collective reading. He does not explicitly state the reading with the singular prefix, but it is implied that it is distributive. (15) shows the contrast for a class II noun, (16) for a class III noun.

(15) (Yumitani 1998:105)
a. *té:hete* *Ø-tʸí*
shirt **3S-**fall.PL
'Shirts (PL) fell off.'

b. *té:hete* ***il-****tʸí*
shirt **3P-**fall.PL
'Shirts (PL) fell off (collectively).'

(16) (Yumitani 1998:105)
a. *tʸê:tiba* **Ø-***tʸí*
box **3S-**fall.PL
'Boxes (PL) fell off.'

b. *tʸê:tiba* ***il-****tʸí*
box **3P-**fall.PL
'Boxes (PL) fell off (collectively).'

Note that these data contrast with the Kiowa data (as in (4), (6) and (7) above) in that the plural rather than the singular is denoting collectivity.[12] Yumitani (1998) also states that a few class I nouns may occur with the plural prefix, his examples being *pʼǽ:* 'moon', *wǫ́hǫ:* 'pumpkin', and *wǫ́hǫ* 'star', all inanimates. The inverse may apparently be used instead, but no example is given, leaving it unclear as to when one is used as opposed to the other.

Table 2 summarizes the pattern of the inverse suffix and pronominal prefixes in the noun class system for Towa.

[12] Given the counter-intuitiveness of this description, this statement should be checked. Unfortunately Yumitani gives only a few examples outside of context, so at present this statement must stand as is.

Table 2. Towa Noun Classes

	Inv Sfx	Index	Inv Sfx	Index	Inv Sfx	Index
Class I		*Ø-*	*-š*	*ɻil-*	*-š*	*e-*
Class II	*-š*	*e-*	*-š*	*ɻil-*		*Ø-/ɨl-*
Class III		*Ø-*	*-š*	*ɻil-*		*Ø-/ɨl-*
Class IV		*Ø-*		*Ø-*		*Ø-*

The semantic basis for noun classes is again not immediately apparent, but for a couple of features. All animates fall under class I along with a handful of inanimates, although far fewer than in Kiowa (e.g. *wɑ́de'li* 'egg', *p'é:* 'moon', *wéhí* 'skeleton'). Class IV is made up entirely of mass/non-count nouns (e.g. *ɻí* 'blood', *wóho'le* 'dough', *sɻí:* 'lard') and shows no sub-classification as Kiowa does. All other inanimate nouns are divided more or less equally between classes II and III (neither or which appears to be a closed class) with no obvious semantic criteria to distinguish them (see section 3.4).

While its morphological realization is very different, the system of Tewa discussed in the next section is almost identical to the Towa pattern presented here.

2.3. TEWA. Tewa (the Pueblos of Nanbé and Ohkay Owingeh—formerly known as San Juan, Pojoaque, San Ildefonso, Santa Clara, and Tesuque, New Mexico, and Tewa Village, Arizona) is spoken at the largest number of Pueblos of any language in New Mexico and as such has several dialect distinctions. With the exception of the Arizona Tewa dialect, which is fairly divergent and perhaps best treated as a separate, albeit very closely related, language for descriptive purposes, the differences among the Rio Grande Tewa dialects are fairly minimal. Unlike Towa, Tewa is not being widely transmitted to the next generation, although several of the Pueblos have some kind of community-based revitalization or teaching program underway. Since the formal marking of noun class is relatively subtle in the language, a linguistic consultant who works with Tewa in one of these communities needs to be aware of the system in order to produce an accurate grammatical description, dictionary, and/or pedagogical materials.

Among the Tanoan languages, Tewa has the largest amount of lexical material available in print; there is very little mention of noun class. Only three sources (A. Speirs 1974; R. Speirs 1966, 1972) explicitly describe the phenomenon,[13] and these state the noun classes for only an extremely limited set of items. The primary source for this omission seems to lie in the limited contexts in which the noun class system overtly manifests itself. There is (almost) no reflection of it in the pronominal indexation on verbs and very few nouns show any overt number marking. It is only by certain modifiers that the organization of the

[13] Harrington (1916), Henderson and Harrington (1914), and Robbins, Harrington, and Freire-Marreco (1916) do make reference to the different "genders" in Tewa, but do not go into any details of the grammar.

system can be seen. The inverse suffix is essentially *-n*, but there are a few lexically determined allomorphs. On nouns, it occurs with only a limited number of class I animates and deverbal animate nouns with relativizer *-i',* but it is still largely productive on adjectives and functional modifiers such as *i'/in* 'the' and *wí/wên* 'a, some'.

The distribution of number marking appears to be identical to that present in Towa: class I is singular in the basic, dual and plural in the inverse. This can be seen reflected in the number agreement on the adjective and demonstrative in the following sentences ((17)-(19) from Speirs 1972:482).

(17) a. *o'i* *wî'* *tsé* ***he:'i*** *na-ke't'ą́*
this.**BAS** one dog big.SG.**BAS** 3S-fall.SG/DU
'This one big dog fell.'

b. *o'i-**n*** *wíye* *tsé* ***he'ennin*** *da-ke't'ą́*
this-**INV** two dog big.**INV** 3D-all.SG/DU
'These two big dogs fell.'

c. *o'i-**n*** *po:ye* *tsé* ***he'ennin*** *di-yemu*
this-**INV** three dog big.**INV** 3P-fall.PL
'These three big dogs fell.'

Class II is plural in the basic, and dual and singular in the inverse. The demonstrative and adjective again agree in number (18), as does the numeral 'one', which takes the form *wéhpí:* in the inverse.

(18) a. *o'i-**n*** ***wéhpí:*** *te:* ***he:'in*** *na-ke't'ą́*
this-**INV** one.**INV** tree big.**SG.INV** 3INAN-fall.SG/DU
'This one big tree fell.'

b. *o'i-**n*** *wíye* *te:* ***he'ennin*** *da-ke't'ą́*
this-**INV** two tree big.**INV** 3D-fall.SG/DU
'These two big trees fell.'

c. *o'i* *po:ye* *te:* ***he'endi*** *na-yemu*
this.**BAS** three tree big.PL.**BAS** 3INAN-fall.PL
'These three big trees fell.'

Class III is singular and plural in the basic and dual in the inverse, as seen in (19).

(19) a. *o'i* *wî'* *k'u:* ***he:'i*** *na-ke't'ą́*
this.**BAS** one rock big.**BAS** 3INAN-fall.SG/DU
'This one big rock fell.'

b. *o'i* ***-n*** *wíye k'u: he'ennin da-ke't'ą́*
this-**INV** two rock big.**INV** 3D-fall.SG/DU
'These two big rocks fell.'

c. *o'i po:ye k'u:* ***he'endi*** *na-yemu*
this.**BAS** three rock big.PL.**BAS** 3INAN-fall.PL
'These three big rocks fell.'

Class IV takes no inverse marker. Whatever the number marking otherwise in a clause, a class IV noun like *ą́'i* 'sugar' in the following sentences will take basic number. The noun can be construed as a collective whole, as in (20a), or as individuated units, as in (20b), reflected in the number suppletion of the stem. In neither case does an inverse marker appear.

(20) (Speirs 1974:59)

a. ***wí*** *ą́'i na-k'ó:*
INDF.BAS sugar 3INAN-be.lying.SG/DU
'There is some sugar (spilled).'

b. ***wí*** *ą́'i na-k^{w}'ǫ́*
INDF.BAS sugar 3INAN-be.lying.PL
'There is some sugar (spilled in clumps).'

Table 3, adapted from Speirs (1974:46), summarizes the noun classes using the modifier *wí*/*wên* 'a, some'.

TABLE 3. Tewa Noun Classes

	Singular	Dual	Plural
Class I	*wí*	*wên*	*wên*
Class II	*wên*	*wên*	*wí*
Class III	*wí*	*wên*	*wí*
Class IV	*wí*	*wí*	*wí*

As in Towa dual is always marked with the inverse (except in class IV where no number distinctions are being made anyway), class I is inverse in the plural, and class II is inverse in the singular. Class III marks both the singular and the plural as basic.

The number of nouns for which class is known is extremely limited; as such it is difficult to determine any semantic basis for the classes and Tewa will not enter much into the comparative discussion in section 3. It is apparent from the little data available that class I is composed primarily—if not almost wholly—of animates. Class IV also appears to consist largely of non-count and abstract nouns (e.g. *p'o:* 'water', *tsonxu:* 'commandment').

This leaves classes II and III to handle most of the inanimate nouns, just as in Towa. Speirs (1974) does arrive at some conclusions about the distribution of nouns in these two classes: 'bulky' objects go in class III (e.g. *k'u:* 'stone', *po:* 'pumpkin', *púwéré* 'chair'); long and flat objects, rooted/attached and long objects, containers, and rooted and bulky objects fall under class II (e.g. *sú* 'arrow', *p'o'k'e:* 'river', *te:* 'tree') (p. 62-63), although it is difficult to determine how pervasive these generalizations really are.

2.4. NORTHERN AND SOUTHERN TIWA. Tiwa is perhaps the Tanoan language most familiar to linguists not working in the Southwest, thanks to a sizable body of literature, including Harrington's detailed (1910) grammatical sketch of Taos Tiwa, George Trager's numerous articles on the same dialect (1936, 1946, 1948, 1954, 1960, 1961, among others), and several articles on Isletan Southern Tiwa by Summer Institute of Linguistics fieldworkers and Donald Frantz in the 1970s through 1990 (Allen et al. 1984 and Allen et al. 1990 inter alia). All of this activity should not indicate that the general policies of the Tiwa Pueblos differ significantly from those of the others. Trager had to work with his three main consultants outside of Taos Pueblo and on the condition of maintaining their anonymity because the research did not have community approval. Even though members of a speech community may desire the assistance of an outside linguist against the policies of the tribal government, it is not for the linguist to instigate the research without the support of some element of the community. The present study obviously benefits from Trager's work, but this kind of method is generally frowned upon by today's fieldworker.

Despite the preceding indictment of the methods of some previous researchers, the results of these studies show that the noun class and number systems of the Tiwa languages are strikingly different than the basic-inverse strategies found among the other Kiowa-Tanoan members, but are still reminiscent enough to allow comparison. The Tiwa languages themselves—Northern Tiwa (Pueblos of Picurís[14] and Taos, New Mexico) and Southern Tiwa (Pueblos of Isleta and Sandía, New Mexico; also, formerly Ysleta del Sur, Texas)—differ in many respects in the manifestations of their noun class systems, but on the whole pattern very similarly, at least superficially (as described for Northern Tiwa in F. Trager 1975, G. Trager 1946, 1961, Zaharlick 1975, 1977, and Southern Tiwa in Allen et al. 1990, Leap 1970a, b, among others). As they are the best described varieties, this paper will focus on the Southern Tiwa spoken at Isleta Pueblo and the Northern Tiwa spoken at Taos.

Like in Kiowa and Towa, the noun class division is best analyzed through the pronominal indexation on verbs. These prefixes demonstrate a formal three-way number/class distinction,[15] labeled A, B, C in the Tiwa literature. The noun classes in the languages,

[14] The Picurís dialect is phonologically divergent enough from the Taos dialect to impede mutual intelligibility. It may be ultimately best, therefore, to treat Taos and Picurís as two separate languages for descriptive purposes. Noun class is one area of the grammars where the languages do seem to differ to a small extent.

[15] There is another indexation class called L which takes a prefix *na-* on the noun stem. This is not reflected as a distinct class in the pronominal prefix paradigm. It is unclear at present how this class interacts with the others, and so it will be set aside for now.

which are labeled I, II, III, are then based upon pair sets of these forms—specifically A-B, B-C, A-C—in which one is singular and one is non-singular. These noun class distinctions are better illustrated by transitive than intransitive pronominal prefixes. The following sentences exemplify this using a constant third singular animate (A) subject while varying the noun class of the direct object. Class I takes A indexation when singular and B indexation when plural, as in (21).

(21) (adapted from Allen et al. 1984:295)

a. *łiawrade* ***Ø-****səan-mų-ban*
woman A:**A-**man-see-PST
'The woman saw the man.'

b. *łiawrade* ***i-****səan-mų-ban*
woman A:**B-**man-see-PST
'The woman saw the men.'

Class II takes B when singular—note the same prefix in (22a) as in (21b)—and C when plural as in (22b).

(22) (Allen et al. 1990:327)

a. *səanide* ***i-****kahun-mų-ban*
man A:**B-**box-see-PST
'The man saw the box.'

b. *səanide* ***u-****kahun-mų-ban*
man A:**C-**box-see-PST
'The man saw the boxes.'

Class III nouns are indexed A when singular—compare (23a) and (21a)—and C when plural—compare (23b) and (22b).

(23) (Allen et al. 1990:327)

a. *səanide* ***Ø-****natufu-mų-ban*
man A:**A-**letter-see-PST
'The man saw the letter.'

b. *səanide* ***u-****natufu-mų-ban*
man A:**C-**letter-see-PST
'The man saw the letters.'

There is also a fourth class of nouns that make no number distinction; these appear to be indexed as C in Northern Tiwa and mostly as A in Southern Tiwa, although there do seem to be some that take either C or B indexation.

Noun class is also reflected, at least in part, on noun stems. In Southern Tiwa this marking is largely limited to class I nouns, with -(*V*)*de* suffixed for singular (A) and -(*ni*)*n* suffixed for non-singular (B),[16] e.g. *səan**ide*** ‘man’, *səan**nin*** ‘men’; *musa**de*** ‘cat’, *mus**an*** ‘cat’. Taos Northern Tiwa, on the other hand, has an overt suffix on all non-incorporated nouns,[17] reflecting the class parallel to the pronominal indexation. The A suffix is *-na*; the B suffixes are unpredictably *-nạ* or *-nemạ*;[18] the C suffix is *-ne*.[19] Examples of the noun class suffixes in Taos Northern Tiwa are given in (24) (data from Trager 1946).

(24)		Singular	Plural	
	I	*łuł’i**na***	*łuł’i**nemạ***	‘old man’
		*tuculo**na***	*tuculo**nạ***	‘hummingbird’
	II	*p’iane**nemạ***	*p’iane**ne***	‘mountain’
		*họlu**nạ***	*họlu’u**ne***	‘lung’
	III	*kwo**na***	*kwo**ne***	‘ax’
	IV	*pha’a**ne***		‘fire’

Tables 4 and 5 present summaries of the (Taos) Northern and (Isleta) Southern Tiwa noun class systems respectively, the pronominal indexation prefixes representing transitive third person singular subject forms with noun class distinctions for the direct object. Of significance for comparison to other Kiowa-Tanoan languages is the fact that the plural indexation of class I is the same as the singular indexation of class II in both Tiwa languages, a pattern in common with the inverse marking in classes I and II of the other languages.

---

[16] Some non-class I nouns may apparently take this suffix under some circumstances (cf. Rosen 1990: 699, fn. 16).

[17] Tanoan languages, the Tiwa languages especially, show a very productive noun incorporation process.

[18] Kontak and Kunkel (1987) describe these allomorphs as being phonologically conditioned, although their analysis does not fit all of the data given by Trager (1946). This needs to be checked further.

[19] Picuris Northern Tiwa has *-ne* for singular A, *-mộ* for singular B, and *-nệ* for plural (B/C).

Table 4. Taos Northern Tiwa

	Singular		Dual		Plural	
	Class Sfx	Index	Class Sfx	Index	Class Sfx	Index
Class I	*-na*	*Ø-*	*-nạ/-nemạ*	*i-*	*-nạ/-nema*	*i-*
Class II	*-nạ/-nemạ*	*i-*	*-ne*	*u-*	*-ne*	*u-*
Class III	*-na*	*Ø-*	*-ne*	*u-*	*-ne*	*u-*
Class IV					*-ne*	*u-*

Table 5. Isleta Southern Tiwa

	Singular		Dual		Plural	
	Class Sfx	Index	Class Sfx	Index	Class Sfx	Index
Class I	*-(V)de*	*Ø-*	*-(ni)n*	*i-*	*-(ni)n*	*i-*
Class II		*i-*		*u-*		*u-*
Class III		*Ø-*		*u-*		*u-*
Class IV						*Ø-/i-/u-*

As in all of the other Kiowa-Tanoan languages, class I for both Northern and Southern Tiwa consists of all of the animate nouns and very few inanimate nouns. Class IV, which makes no distinction for number, contains mostly non-count nouns. It is again classes II and III that divide up most nouns with inanimate referents on a seemingly arbitrary basis.

In an attempt to analyze the semantic organization of the Tiwa noun class systems, Trager (1961) and Leap (1970a, b) describe the A, B, C class marking distinction in terms of a purely number construal schema wherein A is used for a 'unit', B is used for a 'set', and C is used for an 'aggregate'. These terms are only roughly defined in the Tiwa literature. Leap (1970a:202) briefly describes units as "occurring singularly as discrete objects", sets as "composed of items in systematic, patterned relationship to each other", and aggregates as "a simultaneous occurrence of several single items, without any necessary relationship or connection." Nouns then appear in one of these classes depending upon their construal in a given context which roughly, in translation, corresponds to an English singular or plural.

While this innovative analysis may not be too inaccurate diachronically, it suffers from being too *post hoc* in a synchronic analysis. There is no justification as to why a given noun is construed in the way that is represented by the indexation classes with which it appears. Also, if the system were purely based on number construal, one would expect more inanimate nouns in class I (i.e., nouns that may be construed as units (A) or sets (B)) and moreover, perhaps nouns that could take any of the three indexation classes, depending upon the semantic/pragmatic context. The classes appear to be very fixed with very blatant

animate-inanimate and count/non-count distinctions, and not subject to any more construal than singular versus non-singular. Although a system akin to the unit-set-aggregate distinction may be at the core of the Tiwa (and, moreover, Kiowa-Tanoan) noun class division, it does not appear to be a synchronically active system and interactions with other semantic factors render such an analysis unwieldy.

3. COMPARATIVE ANALYSIS. In attempting to reconstruct what the semantic characteristics of the number/noun-class system of Proto-Kiowa-Tanoan might have been, it is necessary to reconcile the differences in the synchronic daughter languages. In their number-marking strategies, two major systems have been presented: the basic-inverse system of Kiowa, Towa, and Tewa, and the 'overlapping' class system of Northern and Southern Tiwa. Furthermore, the basic-inverse strategies also show a division between the Kiowa system, in which the dual is always basic, and the Tanoan (Towa and Tewa) system, in which the dual is always inverse, with the other numbers showing different strategies around this central fact.[20] This section will discuss the motivations for deciding which of these is more representative of how the proto-system may have patterned before further delving into the deeper semantic basis of that pattern.

3.1. KIOWA VERSUS TANOAN INVERSE. A comparison of the Kiowa and Tanoan inverse systems reveals striking similarities but an even more striking difference. The patterns are almost identical but for the status of the dual, it being inverse in every class in Towa and Tewa, but basic in every class in Kiowa. The curiosity comes by the fact that class III receives a corresponding 'reversal of polarity' between the two different systems. These correspondences and differences are even more notable given that the inverse markers in Kiowa, representing them with {*-gɔ́*}, do not appear to bear any historical relation to the inverse markers in Towa, *-š,* or Tewa, *-n*,[21] and yet both sides of the family have maintained an inverse system, without recourse to a less typologically-marked singular/non-singular system.

The basic similarities between the two systems are apparent and may be taken as reflective of the historical proto-number system: there are four classes; the dual always has the same status—basic or inverse—irrespective of class;[22] there is a class that is basic in the singular and inverse in the plural (class I); there is a class that is basic in the plural and inverse in the singular (class II); there is a class the polarity of which in the singular and plural is the same and is opposite that found in the dual (class III); there is a class that does

[20] This is not to suggest a historical causal relation between the (re)categorization of dual number and the semantic or formal organization of the noun class systems.

[21] This is based on impression only. Although scholars have made progress in reconstructing word and stem initial consonants for Proto-Kiowa-Tanoan, vowels and morpheme internal and final consonants have not received much attention (Hale 1962, 1967, Watkins 1977, 1978).

[22] Where there is number-marking, at least. Class IV makes no inverse distinction, and in the Tanoan system, does not even reflect number at all.

not reflect the inverse (class IV); and, indexation reflects four morphologically distinct numbers: singular, dual, plural, inverse. The decision lies in whether the historical dual was basic or inverse and what the semantic distribution of nouns was like in the historical system compared to the synchronic ones. This paper takes the Tanoan pattern to be the more conservative one, Kiowa having innovated in its noun class system.

The first point to note in support of this is that although Kiowa has four noun classes, class III is extremely restricted, apparently containing less than half a dozen members. Class III in both branches of the family is marked as having the dual in opposition to the other two numbers; in Kiowa the dual is basic, but its few members—fruits and hair—do not demonstrate any obvious basis for being construed as basically dual.[23] Fruit and hair tend to come in large quantity or in individual units, not in pairs. Furthermore, as Watkins (1984:88-9) points out, the basic form of a class III noun—used with a singular indexation on the verb—may be used for a varietal construal to denote a plural occurrence of different varieties of the same kind of item (i.e. different types of apples, different heads of hair, etc.), a usage that does not match the pattern of inverse marking. In both of the sentences in (25), the class III nouns have no inverse marking, but are not making any reference to dual number.

(25) (Watkins 1984:88-89)

a. *álɔ̀:* ***bâ-****bɔ́:*
apple.**BAS** 2P:(-2S):**3**S-bring.IMP
'Bring (2PL) me apples (of different varieties)!'

b. *hɔ́ndé* *ɔ́l* ***bɔ́-****sǫ́:mí:*
what.INDF hair.**BAS** (X):2P:3S-interesting
'What interesting (kinds of) hair you (PL) have.'

Class III in Tanoan, on the other hand, is a large open class in which dual is marked as inverse as opposed to a basic singular and plural. Although a few members of class III do have referents that frequently come in pairs (e.g. Towa *hɑ́é:* 'arm', *hǫ́:* 'leg', *mą̂:tyà* 'thumb', *ĩ́:* 'shoe', *wą̂:te* 'glove'), most tend to appear as units or in quantity (e.g. Towa *pé:* 'heart', *p'ǫ́:* 'road', *gí:č'į́'nį́* 'wheat', *kʸ'â:* 'rock'), in keeping with the unmarked status of singular and plural and the marked status of dual.

Similar evidence for innovation in the Kiowa system comes from class IV nouns. In Tanoan, this class appears to consist only of non-count mass and abstract nouns, consistent with the absence of inverse marking on them. In Kiowa, however, this class is more heterogeneous, consisting not only of non-count nouns, but also many inanimate count nouns. This latter fact is reflected in the division of class IV into three subclasses based on the pronominal prefixes conveying number differentiation; although there is no inverse form for these nouns, there are still number contrasts for singular, dual, and plural. Going even further, a comparison of Tables 1 and 2 above shows a correlation between Kiowa classes IVa and b and Towa class III; in both languages, dual is marked dual, singular as singular,

[23] See fn. 8 above.

and plural as either singular or plural, based upon a collective versus non-collective construal operation. It is only the absence or presence of the inverse that distinguishes these classes across these languages. These facts are suggestive of a merger in Kiowa between the historical class III and IV sets with the loss of the inverse in the former. The opposite could be suggested for Tanoan, that a historical class wherein there was no inverse marking has split into the modern classes III and IV when the dual was reconstrued as inverse, but the homogeneity of the synchronic Tanoan class IV in only consisting of non-count nouns is more suggestive of a semantic-based system, unlike the much more heterogeneous Kiowa class IV.

The final argument for innovation in the Kiowa noun classes to be presented here is based upon the animacy gender distinction. In all of the languages, class I contains all animate nouns without exception plus a few inanimate nouns. The difference between Kiowa and Tanoan is the quantity of inanimate nouns in class I: Kiowa demonstrates a larger ratio of class I inanimates than do the other languages. Aside from several objects made by humans, e.g. *t'áyk^{h}ɔ́t^{h}á:* 'scissors', *k'ɔ̀:* 'knife', *t'ɔ́:* 'spoon', *ɔ̀:kɔ̀sɔ́m* 'mirror, window'—all of which may be construed as instruments—many body parts fall under class I. These include several organs that occur in pairs or singularly (Watkins 1984: 82), e.g. *tá:dè* 'eye', *t'èlbǫ̂:* 'knee', *t^{h}én* 'heart', *zǫ́:* 'tooth'. The rest occur in class II, e.g. *gú:* 'rib', *ɔ́lt^{h}ǫ́:* 'head', *mɔ̀:k'ɔ́n* 'nose, beak', *ɔ̀nsó:* 'foot'. In Tanoan, on the other hand, all body parts fall under classes II and III, along with the overwhelming majority of inanimate nouns. Although there is semantic motivation for paired and individual body parts appearing in class I in Kiowa—where the singular and dual are basic—the inconsistency in animacy, which is otherwise so salient in the Kiowa-Tanoan languages, is suggestive of a change from an earlier more consistent system. Although this will not be followed up here, it is notable that for all class III body parts in Towa, where there is information for the noun class of the Kiowa correspondent, the Kiowa body part noun is class I. Compare class III Towa *sé:* 'eye', *pé:* 'heart', *k^{w}ǫ́:* 'tooth' to the class I body parts above. This demonstrates a specific area of the language to be explored for the details of the semantic shift.

The above discussion has compared and contrasted the Kiowa and Tanoan (Towa and Tewa) noun class systems and suggests that it is the Tanoan system that has conserved more features of the original Proto-Kiowa-Tanoan noun classes, whereas Kiowa has innovated numerous aspects. The next section will compare the basic-inverse strategy of Kiowa, Towa, and Tewa to the 'overlapping' noun class system of the Tiwa languages.

3.2. INVERSE VERSUS THE TIWA SYSTEM. Although Kiowa may be historically the most distantly related of the Kiowa-Tanoan languages, it is the Tiwa languages that display the most unique noun class system of the family. While all of the other languages demonstrate a basic form opposed to an inverse form, the patterns differing for three numbers—singular, dual, and plural—across four noun classes, Northern and Southern Tiwa have three marked forms that pattern differently for two numbers[24]—singular and non-

[24] Dual number is distinguished in the pronominal prefixes, but it does not appear to play any significant role in determining noun class membership.

singular—across four noun classes. These two different strategies are not too distinct and are easily reminiscent of one another, but the question arises as to which is the historical precedent. It will take a more in-depth study taking into account formal cognates and grammatical reconstruction to be certain, but it appears that Tiwa has been the more innovative in this area and that Proto-Kiowa-Tanoan has a basic-inverse system similar to that found in Kiowa, Towa, and Tewa.

The first piece of evidence for this conclusion stems from the language family tree. In the supposed linguistic history of the family (cf. Davis 1959, Watkins 1977), Kiowa split off first, then Towa, and lastly Tewa and Tiwa split (with subsequent divisions between Northern and Southern Tiwa). Tewa and Tiwa demonstrate the most salient cognates between any of the language branches, perhaps including the form of the 'inverse' marker, which is *-n* [ŋ] in Tewa and *-(ni)n* in Southern Tiwa (interpreting the plural suffix as the historical inverse). While Tiwa has a unique noun class system, Tewa patterns almost identically to the more distantly related Towa, suggesting that it has retained an older grammatical system while the closely related Tiwa has changed.

Another suggestive feature is the number categories that are distinguished. While all of the Kiowa-Tanoan languages have overt marking for singular, dual, and plural in their indexation systems, Tiwa is the only branch that shows no correspondence for dual in its number/noun-class distinction. Since it seems more likely for one language branch to have lost a distinct and typologically-marked (if common) category (Croft 2003, Greenberg 1966), than for three branches to have developed a distinction where none previously existed, the Towa-Tewa pattern again shows the greater potential for conservativeness. The loss of the dual category in nominal morphology in Tiwa may itself have precipitated a reevaluation of how the noun classes pattern, subsequently eliminating the distinction between the basic numbers and the inverse.

A survey of the pronominal prefixes in Northern and Southern Tiwa would undoubtedly shed even more light upon the relationship between the synchronic noun class/number distinctions and the diachronic, but given the complexity of the Kiowa-Tanoan pronominal prefix systems, this is a matter for future study. One fact of overt morphology that does straightforwardly suggest a greater intricacy within the Tiwa system is the presence of *four* noun class suffixes on nouns in Taos Tiwa where only *three* categorial distinctions are being made. While the A class is marked *-na* and unambiguously denotes singular number and the C class is marked *-ne* and unambiguously marks plural number, the B class marker has two lexically determined allomorphs, *-nạ* and *-nemạ*, and denotes singular or plural, depending upon with which noun class—I or II—it is occurring. The usage of these two allomorphs does not appear to be determined phonologically[25] or morphosyntactically, but rather seems to be determined by the stem to which it attaches, whether the morpheme is indicating singular or plural. These facts loosely suggest that there may have formerly been a distinction between *-nạ* and *-nemạ* that has since been lost, namely that one of them may have been the dual marker and one the inverse (assuming that the A and C markers have retained their respective singular and plural functions from the historical system). A four way number distinction—singular, dual, plural, and inverse—is exactly what is found in

[25] But see fn. 17 above.

Towa and Kiowa pronominal indexation, and, per the Northern Tiwa facts, may have formerly been distinguished on nouns as well. This is at present highly speculative and more reconstruction of sound correspondences and morphology is needed before this hypothesis can be further elaborated, whether affirmatively or negatively.

3.3. SUMMARY OF KIOWA-TANOAN NOUN-CLASS MORPHOSYNTAX. The preceding discussion indicates that the Proto-Kiowa-Tanoan number and noun class system may have had the pattern found in Table 6.

TABLE 6. Proto-Kiowa-Tanoan Number and Noun Class Marking

	Singular		Dual		Plural	
	N Sfx	Index	N Sfx	Index	N Sfx	Index
Class I	(SG)	SG	DU/INV	DU	INV	INV
Class II	INV	INV	DU/INV	DU	(PL)	SG/PL
Class III	(SG)	SG	DU/INV	DU	(PL)	SG/PL
Class IV					(PL)	PL

Those categories in parentheses may or may not have been formally marked. There are three numbers distinguished, with dual consistently being demarcated as dual, whether or not it is formally inverse (since it is consistently distinguished from the inverse in pronominal indexation). Class I marks singular as singular, but considers the plural to be inverse. Class II marks the singular as inverse, and considers the plural to be basic, using singular or plural indexation depending upon the construal of number as either collective or distributive respectively.[26] Class III considers both the singular and the plural to be basic, again marking the plural as either singular or plural depending upon construal. Class IV is limited to non-count nouns, which it formally marks as plural, although construal operations may apply which might modify this.

What the exact semantic nature of the inverse number category may have been remains unresolved. It contrasts with overt categorization for the same numbers that it conveys, so the motivation for its presence is unclear. The collective versus distributive construal is already made in some of the synchronic languages by the singular-plural distinction when the referent is plural, so there is no indication that the inverse might perform or might have performed such a construal function. In part, determining its semantic basis depends upon determining whether the dual is or was truly considered inverse, given that it has its own distinct marking in indexation. It also depends upon determining the semantic basis for the relegation of nouns to the noun classes; in doing so, the similarity between the plural of class I and the singular of class II might be analyzed, and the function of the inverse

[26] It is unclear whether this feature was part of the Proto-KT system or if it was independently developed in Kiowa and Towa.

revealed to a degree. The following section will discuss what has been found concerning the semantic basis for the categories. However, since there are still open questions about the distribution of nouns across noun classes, the historical semantics of the inverse will remain unresolved for the present.

It must be borne in mind that the inverse system did not necessarily have a transparent semantic basis in Proto-Kiowa-Tanoan.[27] Proto-languages are of course still languages, unattested though they may be, and are just as likely to show arbitrary properties as synchronic languages. Arbitrary as grammars may end up being, however, they usually do not start out that way. Constructions and patterns tend to have an original semantic or formal motivation which is then obscured over time by grammaticalization processes. The questions driving this paper are targeted towards these original motivations, even though they may no longer have been apparent by the time of the immediate ancestor of the modern Kiowa-Tanoan languages. That such a typologically unusual inverse number system is not found in any other language family—including Uto-Aztecan, which some have posited has a connection to Kiowa-Tanoan (Whorf and Trager 1937, Davis 1989)—suggests that the system and the motivations for the system arose after an ancestor of the Kiowa-Tanoan languages had split from other relations. Only further research will be able to determine the possible character of Proto-Kiowa-Tanoan and whether the inverse system had any transparent semantic basis in that language or if it had been lost by this time.

3**.4. GENDER AND NUMBER IN HISTORICAL KIOWA-TANOAN NOUN CLASSES.** Why nouns fall into the classes that they do in each of the languages is not entirely clear. If it were based solely upon fine number differentiations, one would expect more variation in the number marking of a given noun, depending upon the construal necessary in the given context, as discussed at the end of section 2.4. Additionally, there would not be the semantic patterns that are quite prevalent in the languages. The most pervasive of these patterns is the restriction of animate nouns to class I and, moreover, the restriction of class I to animate nouns, with very few exceptions. Even in Kiowa, where more inanimates are classed in I than in the other languages, they still are relatively few in number. This provides strong evidence that the languages are at the very least making an animate-inanimate distinction and that this distinction is reflected in the number marking pattern.[28] The question then arises whether there is a further semantic basis between inanimate nouns in their respective classes, i.e., II, III, and IV.

Class IV nouns, which appear in all languages but Kiowa to consist entirely of inanimate non-count and abstract nouns, are obviously distinguished for their unique number properties, namely that no number distinction tends to be made. This is a class determined solely by number semantics and does not seem to have any interaction with any gender distinctions, outside of the fact that most or all non-count and abstract nouns are inanimate.

---

[27] Thanks to an anonymous reviewer for reminding the author to explicate this point.

[28] The distinction between animates and inanimates is well attested elsewhere in the languages. See especially Allen, Gardiner, and Frantz (1984), Rosen (1990), and Sadock (1985) for Southern Tiwa; Yumitani (1998) for Towa.

It is in classes II and III where the semantic basis is most unclear. Each of the languages shows a different pattern for these two classes, which otherwise do seem to be comparable, at least among the Tanoan languages. For instance, all trees appear to fall under class II in both Towa and Northern Tiwa (as does *te:* 'tree' in Tewa), while in Northern Tiwa, at least, the fruits and parts of the plants largely seem to be class III. Conversely, in Southern Tiwa, trees appear to fall mostly under class III.[29] Such patterns are present, but not with enough consistency to make any strong statements without a more detailed analysis. Among the body part terms, which all occur in classes II and III in the Tanoan languages, many class II nouns in Taos Tiwa have a semantic correspondence to class III nouns in Towa and many class III nouns correspond to class II, but there are exceptions (data from Trager 1946 and Yumitani 1998).

(26)

Taos		Towa		
xonemą	(II)	*hé:*	(III)	'arm'
tsinemą	(II)	*sé:*	(III)	'eye'
phona	(III)	*ɸó:lá*	(II)	'hair'
phəyna	(III)	*ɸôse*	(II)	'nose'

As discussed in section 2.3, Speirs (1974) points out a distinction between classes II and III in terms of object shape, which is consistent with the Towa and Taos distinction among (tall, thin) trees and their (usually roundish) parts, but this simple categorization would need to be greatly extended to encompass all inanimate count nouns in the languages.

This discussion has ignored the effects of the innovations in the different synchronic systems in trying to determine the historical basis for the two classes. That the closely related Northern Tiwa and Southern Tiwa should assign nouns of similar semantics (and often cognate forms) to different classes indicates that the diachronic change in the noun class system may have far reaching effects, despite a similar underlying semantic basis that may still be residually present in the languages. By the same token, the historical distinction between classes II and III may have been purely based on number construal, with other semantic patterns arising by virtue of the fact that nouns in the same or similar semantic fields will be construed for number in similar ways.

If the basis for the distinction between classes II and III is number, a return to Table 6 would be fruitful. Classes II and III are only distinct in the singular, whether it is marked as singular or as inverse; dual is consistent between them and plural is largely subject to construal between a collective and distributive reading. The difference appears, therefore, to be determined by the semantics of the inverse. Class III nouns make a simple singular, dual, plural distinction where singular is essentially unmarked, i.e., it is not 'unusual'—assuming a direct meaning form correlation—to construe a class III noun as singular. Class II nouns, on the other hand, may occur regularly as dual or plural, but have to be 'specially' marked if one is to construe them as singular. As stated above, the semantic attributes of

[29] Further data are needed to clarify this point.

the 'unusual' and 'special' inverse are unclear at present, but it is appearing more and more that it may be at the heart of the class distinctions.[30]

4. CONCLUSION. This paper has compared and contrasted the number and noun class systems of the major branches of the Kiowa-Tanoan language family. It has suggested that the historical pattern from which all of the synchronic systems arise is most similar in organization to that found in Towa and Tewa, but may not have classified the dual as categorically inverse. This study has also confirmed that there was likely at least an animate-inanimate gender distinction and has suggested that the further division of noun classes is determined primarily or wholly by the distinctions in number. More research is necessary in order to more accurately determine the features around which the historical and modern noun class systems are based. This research will especially need to focus on the reflection of noun class and number in all of the pronominal prefix paradigms, and the correlation of noun classes when taking formal cognates—as well as semantic correspondences—into account. Furthermore, the semantics of the inverse morpheme needs to be isolated in both synchronic and diachronic perspective before there can be a clear picture of the Kiowa-Tanoan noun class system.

Without a doubt the input of native speakers of Kiowa-Tanoan languages would be of great assistance in analyzing the synchronic noun class systems in order to reconstruct the ancestral form. In addition to being able to provide the noun classes of lexical items and narratives to illustrate the nouns in context, they could also give insight into their own subjective semantic construal of the nouns in context. Are class III inanimate nouns construed different than class II nouns when it comes to number? On what basis is a newly introduced lexical item assigned to a given class? To these and more questions fieldwork might be able to provide some answers. The restrictions imposed upon outside linguists doing research on languages in the Pueblos and in similar social situations do raise the serious question of how to feasibly undertake this fieldwork without impinging upon the wishes of the native speakers. Assuming speakers even agree to work with the linguist, possibilities include: doing the research without publishing the results (except in any community-internal language reference materials); publishing the metadata and analysis without including any actual language data (or, exemplifying only with language data that is already publicly available, even if the analysis is actually based on unpublished examples); publishing the data with the analysis, but encrypting it, a possibility that has not been actualized in any publications to the present author's knowledge (although see Debenport 2009, in which the language data are present, but obfuscated in the published version, leaving only the English translations visible). The specific approach taken will be at the whim of the individuals and institutions involved, but options do exist.

It must also be remembered that modern-day fieldwork is not just about the linguist's questions, but about the consultants' too. Native speakers have their own goals regarding

[30] An anonymous reviewer brought up the question of the language contact situation in Proto-Kiowa-Tanoan times and hence, suggesting the possibility of a borrowed system overlying the native system of noun categorization. For purposes of space, this intriguing point cannot be pursued here, but should be kept in mind for future research.

their language and linguistic work and moreover, as the possessors of the language and of the responsibility for its continued use and transmission, have the right to state what kind of work can be done. The outside linguist conceives of his or her research as contributing to the greater knowledge of how a given language—and language in general—functions. In doing fieldwork and working with native speakers, it is however his or her task to put that knowledge to use by sharing it with the people actually speaking the language and making sure their goals are met. Typically these goals revolve around maintaining and revitalizing the language within the community by assisting in the development of language classes and pedagogical materials. This may be accompanied by the caveat that the data involved in such work must be kept confidential, within the community. This may mean that the linguist's own research questions are sidelined or put under restrictions (e.g. unable to publish) in respecting the desires of the community. Ideally, there would be native speakers with the training to do linguistic research on their own languages who would also have the community-internal knowledge and experience of how to manage language data without compromising community ideology. However, there are presently no fully trained linguistic researchers now living known to the author who are native speakers of a Tanoan language.[31] Although the above situations are not unique to the Southwest or the Kiowa-Tanoan language family, this discussion should serve as a poignant reminder of both the realities of actually accomplishing the fieldwork needed to further linguistic research as well as of the responsibilities of the linguist towards the people he or she is researching.

[31] However, Gus Palmer, Jr., a speaker of Kiowa and a member of that community, is an associate professor at the University of Oklahoma with a focus in linguistic anthropology.

References

Allen, Barbara J., Donna B. Gardiner & Donald G. Frantz. 1984. Noun incorporation in Southern Tiwa. *International Journal of American Linguistics* 50(3). 292-311.

Allen, Barbara J., Donald G. Frantz, Donna B. Gardiner & David M. Perlmutter. 1990. Verb agreement, possessor ascension, and multistratal representation in Southern Tiwa. In Paul Postal & Brian Joseph (eds.), *Studies in relational grammar 3*, 321-383. Chicago: University of Chicago Press.

Croft, William. 2003. *Typology and universals,* 2nd edn. Cambridge: Cambridge University Press.

Davis, Irvine. 1959. Linguistic clues to northern Rio Grande prehistory. *El Palacio* 66. 73-84.

Davis, Irvine. 1989. A new look at Aztec-Tanoan. In Mary Ritchie Key & Henry M. Hoenigswald (eds.), *General and Amerindian ethnolinguistics: In remembrance of Stanley Newman*, 365-379. New York: Mouton de Gruyter.

Debenport, Erin. 2009. *"Listen so you can live life the way it's supposed to be lived": Paradoxes of text, secrecy and language at a New Mexico Pueblo*. Chicago, IL: University of Chicago dissertation.

Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of grammar,* 2nd edn., 73-113. Cambridge, MA: MIT Press.

Hale, Kenneth. 1962. Jemez and Kiowa correspondences in reference to Kiowa-Tanoan. *International Journal of American Linguistics* 28(1). 1-5.

Hale, Kenneth. 1967. Toward a reconstruction of Kiowa-Tanoan phonology. *International Journal of American Linguistics* 33(2). 112-120.

Harrington, John P. 1910. An introductory paper in the Tiwa language, dialect of Taos, New Mexico. *American Anthropologist* 12(1). 11-48.

Harrington, John P. 1916. *The ethnogeography of the Tewa Indians* (Bureau of American Ethnology Annual Report 29). Washington, DC: Smithsonian Institution.

Harrington, John P. 1928. *Vocabulary of the Kiowa language* (Bureau of American Ethnology Bulletin 84). Washington, DC: Smithsonian Institution.

Henderson, Junius & John P. Harrington. 1914. *Ethnozoology of the Tewa Indians* (Bureau of American Ethnology Bulletin 56). Washington, DC: Smithsonian Institution.

Kontak, Corrie & Janet Kunkel. 1987. Grammar sketch of Northern Tiwa, Taos dialect. Unpublished ms.

Leap, William Lester. 1970a. *The language of Isleta, New Mexico*. Dallas, TX: Southern Methodist University dissertation.

Leap, William L. 1970b. Tiwa noun class semology: A historical view. *Anthropological Linguistics* 12(2). 38-45.

Merrifield, William R. 1959. Classification of Kiowa Nouns. *International Journal of American Linguistics* 25(4). 269-271.

Mooney, James. 1898. *Calendar history of the Kiowa Indians* (Bureau of American Ethnology Annual Report 17). Reprinted 1979: Smithsonian Institution Press.

Robbins, Wilfred W., John P. Harrington & Barbara Freire-Marreco. 1916. *Ethnobotany of the Tewa Indians* (Bureau of American Ethnology Bulletin 55). Washington, DC: Smithsonian Institution.

Rosen, Carol. 1990. Rethinking Southern Tiwa: The geometry of a triple-agreement language. *Language* 66(4). 669-713.

Sadock, Jerrold M. 1985. The Southern Tiwa incorporability hierarchy. *International Journal of American Linguistics* 51(4). 568-572.

Speirs, Anna. 1974. Classificatory verb stems in Tewa. *Studies in Linguistics* 24. 45-64.

Speirs, Randall H. 1966. *Some aspects of the structure of Rio Grande Tewa*. Buffalo, NY: State University of New York at Buffalo dissertation.

Speirs, Randall H. 1972. Number in Tewa. In M. Estellie Smith (ed.), *Studies in linguistics in honor of George L. Trager* (Janua Linguarum, series maior 52), 479-486. The Hague: Mouton.

Sprott, Robert W. 1992. *Jemez syntax*. Chicago, IL: University of Chicago dissertation.

Takahashi, Junichi. 1984. *Case marking in Kiowa: A study in the organization of meaning*. New York, NY: City University of New York dissertation.

Trager, Edith Crowell. 1960. *The Kiowa language: A grammatical study*. Philadelphia, PA: University of Pennsylvania dissertation.

Trager, Felicia Harben. 1975. Morphemic change in Picuris: A case of culture contact? *Studies in Linguistics* 25. 89-93.

Trager, George L. 1936. The language of the Pueblo of Taos. *Maître Phonetique* 56. 59-62.

Trager, George L. 1946. An outline of Taos grammar. In Cornelius Osgood (ed.), *Linguistic structures of Native America*, 184-221. New York: Viking Fund.

Trager, George L. 1948. Taos I: A language revisited. *International Journal of American Linguistics* 14(3). 155-160.

Trager, George L. 1954. Taos II: Pronominal reference. *International Journal of American Linguistics* 20(3). 173-180.

Trager, George L. 1960. Taos III: Paralanguage. *Anthropological Linguistics* 2(2). 24-30.

Trager, George L. 1961. Taos IV: Morphemics, syntax, semology in nouns and in pronominal reference. *International Journal of American Linguistics* 27(3). 211-222.

Trager, George L. & Edith Crowell Trager. 1959. Kiowa and Tanoan. *American Anthropologist* 61(6). 1078-1083.

Watkins, Laurel J. 1977. Kiowa-Tanoan internal relationships. In Donald M. Lance & Daniel E. Gulstad (eds.), *The 1977 Mid-American Linguistics Conference*. Columbia, MO: University of Missouri.

Watkins, Laurel J. 1978. On *w and *y in Kiowa-Tanoan. *The Annual Meeting of the Berkeley Linguistics Society* 4. 477-484.

Watkins, Laurel J. 1982. Passives and verb agreement in Kiowa-Tanoan. *Kansas Working Papers in Linguistics* 7. 209-221.

Watkins, Laurel J. 1996. Reconstructing person and voice in Kiowa-Tanoan: Pitfalls and progress. *The Annual Meeting of the Berkeley Linguistics Society* 22. 139-152.

Watkins, Laurel J. with Parker McKenzie. 1984. *A grammar of Kiowa*. Lincoln: University of Nebraska Press.

Whorf, Benjamin L. & George L. Trager. 1937. The relationship of Uto-Aztecan and Tanoan. *American Anthropologist* 39(4). 609-624.
Wonderly, William L., Lorna F. Gibson & Paul L. Kirk. 1954. Number in Kiowa: Nouns, demonstratives, and adjectives. *International Journal of American Linguistics* 20(1). 1-7.
Yumitani, Yukihiro. 1998. *A phonology and morphology of Jemez Towa*. Lawrence, KS: University of Kansas dissertation.
Zaharlick, Amy. 1975. Pronominal reference in Picuris. *Studies in Linguistics* 25. 79-87.
Zaharlick, Amy. 1977. *Picuris syntax*. Washington, DC: American University dissertation.

Logan Sutton
ldsutton@unm.edu

5

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 91-123
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4452

The Story of *ô in the Cariban Family[1, °]

Spike Gildea[a], B. J. Hoff[b], and Sérgio Meira[c]

[a]*University of Oregon*
[b]*Leiden University*
[c]*Koninklijke Nederlandse Akademie van Wetenschappen/Leiden University*

This paper argues for the reconstruction of an unrounded mid central/back vowel *ô to Proto-Cariban. Recent comparative studies of the Cariban family encounter a consistent correspondence of *ə : o : ɨ : e*, tentatively reconstructed as *o_2 (considering only pronouns; Meira 2002) and *ô (considering only seven languages; Meira & Franchetto 2005). The first empirical contribution of this paper is to expand the comparative database to twenty-one modern and two extinct Cariban languages, where the robustness of the correspondence is confirmed. In ten languages, *ô merges with another vowel, either *o or *ɨ. The second empirical contribution of this paper is to more closely analyze one apparent case of attested change from *ô > *o*, as seen in cognate forms from Island Carib and dialectal variation in Kari'nja (Carib of Surinam). Kari'nja words borrowed into Island Carib/Garífuna show a split between rounded and unrounded back vowels: rounded back vowels are reflexes of *o and *u, unrounded back vowels reflexes of *ô and *ɨ. Our analysis of Island Carib phonology was originally developed by Douglas Taylor in the 1960s, supplemented with unpublished Garifuna data collected by Taylor in the 1950s.

1. INTRODUCTION: THE ROLE OF FIELDWORK IN COMPARATIVE WORK. This paper is one in a recent series of steps forward in comparative phonological work in the Cariban family, in this case focusing on the reconstruction of central vowels. While comparative work is not usually the first thing one thinks of when opening a volume on fieldwork, all comparative work is predicated on the ability of the comparativist to compare

[1] We would like to thank the following organizations for support in conducting the fieldwork that produced some of the data cited here: NSF grants No. BNS-8609304, DBS-9210130/BNS-9318847, BNS-9818244, BNS-9909118, and BNS-0117619; the Max-Planck-Institut für Psycholinguistik in Nijmegen, and the Volkswagen-Stiftung (DoBeS initiative). A preliminary version of this paper was presented at the University of Oregon Phonology working group, and at the 2007 Workshop on American Indian Languages; thanks to the participants at both venues for stimulating discussion of the issues involved in reconstructing vowel quality. Thanks also to the anonymous reviewer for helpful comments, and to Andrea Berez for stimulating our discussion of the necessary dependence of comparative work on reliable fieldwork.

° Editors' note: for the sake of legibility, italicization is used in this paper to represent individual graphemes and orthographic forms in the subject languages; plain text is used for proto-forms marked with an asterisk, for phonetic and phonemic representations given between forward slashes or square brackets, and for forms in direct quotes. See also fn. 4.

reliable data, and for most language families of the world, reliable data only exists as a result of good fieldwork. In South America, all early comparative work—and even some relatively modern comparative work—is weakened by the absence of reliable data. Many early materials are simply lists of words and phrases collected by untrained explorers; some apparently could not hear important phonological distinctions in the new languages they encountered and even those who had reasonably good ears did not have good orthographic tools to consistently write down the distinctions they were hearing. This paper illustrates the importance of good modern fieldwork in two ways: first, after collecting reliable modern data, many inconsistencies in cognate sets disappear and previously unseen patterns become clear; second, reliable modern data can provide insight into previously opaque transcription systems used by older sources, thereby enabling us to make better use of language data recorded several hundred years ago.

Languages belonging to the Cariban language family are spoken in northern South America, the bulk across northern Brazil, Venezuela, Guyana, Surinam, and French Guiana, with outliers to the west in Colombia and to the south in Central Brazil. The historical literature on the Cariban family names over 100 languages, maybe half supported by actual linguistic information (e.g., word lists, brief collections of utterances), and of these, just more than half possibly representing distinct languages. Clearly, many of the languages spoken at the time of first contact with Europeans have become extinct, with an unknown number leaving behind no record. Currently, some 25 Cariban languages remain, with a cumulative total of between 60,000 and 100,000 speakers; census figures in individual countries do not distinguish number of speakers separately from number of members of ethnic group. Well over half the speakers belong to the three closely related languages of the Pemóng Proper Subgroup (Kapóng, Makúshi, and Pemón) and Kari'nja (a.k.a. Carib proper, Galibi, Kari'nya, Kaliña, Cariña). The actual number of speakers for the rest of the Cariban languages falls between around 3000 (e.g., Tiriyó) and a handful of elders (e.g., Mapoyo). Several Caribbean languages historically claimed to be Cariban (Black Carib, Island Carib, Garifuna) are linguistically Arawakan, with some Cariban features (mostly vocabulary) due to intensive contact with Kari'nja invaders (Hoff 1995, cf. also section 4). Older classifications of the family (with the exception of Girard (1971)) are clearly flawed; we present our current classification—more conservative, but we hope also more reliable—in section 3.

In section 2, we review previous comparative treatments of mid vowels in the Cariban family, highlighting problems in the data that led to the earlier reconstruction of two instead of three mid vowels. In section 3, we present the improved correspondences that appear when comparing cognates from a modern, reliable lexical database. These correspondences clearly justify reconstruction of a third Proto-Cariban mid-vowel, *ô.[2] In section 4, we show how the combination of modern fieldwork and examination of correspondences

2 Meira (2002) originally proposed *o_2, which is appropriately neutral with regard to phonetic interpretation, but which, as a digraph, is inconvenient to work with in reconstructed morphology and lexicon. Thus, for aesthetic reasons, we adopt Meira & Franchetto's (2005) shift from *o_2 to *ô. The reader who wants to pronounce *ô may jump to section 5, where we offer arguments in favor of a phonetic value of [ə] or [ɤ].

for *ô combine to cast light on a vexing problem with early work on Island Carib/Garifuna, which in turn cements our reconstruction of the third mid vowel. Finally, in section 5 we summarize our conclusions about this third mid vowel, along with a call for more instrumental acoustic analyses of modern field data.

2. PREVIOUS COMPARATIVE WORK ON CARIBAN MID VOWELS. Modern, reliable comparative work on the Cariban family began with the dissertation of Victor Girard in 1971. In the preceding century, several dedicated philologists (especially Adam 1896 and de Goeje 1909/1946) collected extensive databases, but, as expounded by Girard (1971), they suffered from so many methodological flaws that none of their conclusions could be sustained. However, Girard's generally sound methodology and careful collation of data were in turn confounded by the low quality of the sources available to him. In section 2.1, we review Girard's contribution to the reconstruction of mid vowels. In recent years, tremendous advances have been made in the quality of descriptive data available for comparison, especially in the transcription of central vowels. Thanks to these advances, a series of recent papers has advanced the strong hypothesis that a third mid vowel must be reconstructed to Proto-Cariban. In section 2.2, we review these studies.

2.1. PROBLEMS WITH MID VOWELS IN EARLY COMPARATIVE WORK. Early sources of data from Cariban languages were vexed with poor transcription, especially of vowels. As a result, early attempts to compare Cariban wordlists resulted in inconsistent correspondences. In our own fieldwork, we have found that old word lists are rarely confirmed by our modern transcriptions, and that even the more modern sources consistently mistranscribe vowel quality and length. In particular, the central vowels [ɨ] and [ə] are often mistranscribed as front rounded vowels (*ü, ö, y, ø*) diphthongs (*ui, ue*), or collapsed with any one of the existing vowels (especially *e, o, a,* but sometimes also *i, ɨ, u*). In part as a result of this morass of mistranscription, Girard (1971:77) saw evidence for reconstructing only 6 vowels, *i, *ɨ, *u, *e *o, and *a. Girard did make it clear that he recognized some of the limitations of his database: "[t]here was an inability on the part of the grammarians, both native Spanish and Italian speakers, to distinguish between the mid-front e and the high- or mid-central ɨ ... In general ɨ is written <e> or, when occurring after bilabials, particularly p, <ue>" (p. 45). We have seen many examples of such mistranscription in Girard's sources for languages on which we have worked; in this paper, we illustrate such problems by adding cognates for †Tamanaku and †Kumana to the tables in Appendix 2, where the correspondences of *a : ə* and *e : ə* are almost certainly due to mistranscription.

Even though Girard was aware of these problems, especially with the colonial sources, it is also apparent from his prose that he chose to disregard some sources that indicated modern seven-vowel systems, adding to the six a mid-central vowel: "*ɨ appears to have represented a phonemic area rather than a point. Even modern, and one hopes, more accurate phonetic transcriptions show both [ɨ] and [ə] (see Makiritare, Tiriyó, etc.). Some linguists have analyzed these phones as phonemes. There is no evidence however from other sister languages that Proto-Carib contained three central vowels" (p. 78-9). Interestingly, he does not present the problem as one of what to do with the correspondences

containing these "extra" phonemes, but rather he asserts their irrelevance without careful argumentation.

The problem with these mid-central vowels was exacerbated by the problem of synchronic ablaut, in which nouns, verbs, and postpositions present an alternation between stem-initial *e* (the front grade form) and either *o* or *ə* (the back-grade form). Since the unrounded back-grade form *ə* was often mistranscribed, ablaut was not particularly straightforward to see in many of the languages. This meant that Girard primarily saw the half of ablaut that involved the rounded back-grade form: "Alternating *e ~ *o and *e ~ *a. Many languages manifest these alternating vowels when they occur word-initially. This is a vexing problem for a synchronic as well as a diachronic analysis ..." (p. 83). As a result, Girard (1971:83-6) reconstructed an initial alternation to many Proto-Cariban lexical items, and got used to seeing *e* : *o* correspondences as the outcome of a 'vexing' alternation.

Overall, the many cases of *ə* : *o* correspondences in the older data also generally involved either real *e*, or mistranscribed *e, ue, a,* etc., and hence Girard reconstructs them as either the *e ~ *o alternation, as *e, or as *ɨ. But in fact, Girard did not address most of the cognates we will present in this paper, possibly because their vowel correspondences were simply too chaotic given the poorly-transcribed data available to him at that time. His conclusions stood unchallenged for 30 years, while we awaited sufficient descriptive data to allow us to ask the right questions.

2.2. THE CONTRIBUTION OF MODERN DATA AND RECENT COMPARATIVE WORK. Primarily over the last ten years, at least some reliably transcribed data have become available from nearly every extant language in the family. At least some earlier sources for several languages were reliable, but because other unreliable sources existed for each language, comparativists had no way to know which source to trust. Among us, we three have personally collected data for all but a handful of the languages in the family, which has made it possible in most cases to distinguish the useful sources from those that are not usable. Not all reliable sources are equally rich, and not all issues with transcription have been resolved for all of them (especially of prosodic features, such as vowel length), but an unrounded mid non-front vowel is now clearly confirmed as a phoneme (generally transcribed *ə, ë,* or *ö*) in ten modern Cariban languages: Akuriyó (Meira 2000), Bakairi (Meira 2003), Karihona (Meira 2000), Kapóng (Akawaio: Cesar-Fox 2003; Ingarikó: Souza Cruz 2005), Mapoyo/Yabarana/Pémono (Mattéi-Muller 2003), Panare (Mattéi-Muller 1994), Pemón (Taurepán: Álvarez 2000; Arekuna: Edwards 1978), Tiriyó (Meira 1999, 2000; Carlin 2004), Wayana (Tavares 2005), and Ye'kwana/De'kwana (Hall 1988). In light of these modern findings, Jesuit descriptions can be seen to suggest another mid vowel in Tamanaku (Mattéi-Muller & Henley 1990) and Kumaná (Cumanagóta, Chaimá: Mattéi-Muller & Henley 1990).[3] We now know that in older materials, *ə* (rather than *ɨ*) was frequently mistranscribed as *e*, leading to orthographic collapse of a phonemic distinction and confusion in correspondences.

[3] For a more modern case, cf. the mistranscription of Pemón *ə* as *e* or *ue* in Armellada and Salazar's (1981) dictionary.

Based on these modern sources (as well as his own primary fieldwork with nearly half the languages of the family), Meira (2002) re-opened the question of how many mid-vowels to reconstruct to Proto-Cariban when he compared pronoun sets across the family. In comparing these pronouns, two robustly different correspondence sets emerged, unlikely to be due to a conditioning environment. Since both correspondences present *o* in many languages, Meira posited Proto-Cariban *o in opposition to *o_2: *o becomes *o* in every modern language, but *o_2 develops a range of modern reflexes, including *o, ə, a, ɨ,* and *e*. Meira suggests (p. 260, note 7) "it is not impossible that Proto-Cariban *o_2 was actually *ë ([ə])."[4]

But of course, so limited a lexical selection as the pronouns could conceivably present spurious correspondences, given the sorts of idiosyncratic sound change often associated with high frequency lexical items. The correspondence was confirmed in a much broader sample of the lexicon by Meira & Franchetto (2005). They collected cognate sets with reliable modern data from seven languages, showing (p. 168ff) that not only do the correspondences remain consistent beyond the pronoun sets, but they are, in fact, *more* consistent in the rest of the vocabulary (see Table 1).

Table 1. Correspondences from Meira & Franchetto (2005)

PC	Yukpa	Tiriyó	Hixkar-yana	Makushi	Bakairi	Ikpéng	Kuikúru
*o	*o*	*o*	*o*	*o*	*o*	*o*	*o*
*ô	*o*	*ə*	*o*	*ɨ*	*ə*	*o*	*e*

Meira & Franchetto change the symbol *o_2 to *ô, but they otherwise make no changes to Meira's (2002) reconstruction. In considering the likely phonetic value of *ô, they leave open the possibility of various phonetic realizations: "*ô might end up being *ə, or "unstressed *o*," or *o* + a "schwa-coloring laryngeal," or "any other future solutions" (p. 171).

An additional aid to recognizing modern reflexes of *ô comes from recent advances in the understanding of root-initial ablaut. Descriptions of ablaut show *e/o* and *e/ə* alternation conditioned by person-marking conjugation in nouns, verbs, and postpositions, as well as by a valence-reducing prefix in verbs (Meira, Gildea & Hoff 2007). These correspondences are now well-documented, and the various ablaut alternations are now understood as the outcome of various combinations of these historical changes:

[4] In the attempt to avoid confusion, we here note that there are three different phonetic alphabet traditions in play in this paper. The modern IPA values of the vowels in question are [ɨ] (high central unrounded), [ɯ] (high back unrounded), [ə] (mid central unrounded) and [ɤ] (mid back unrounded). Traditional Americanist IPA symbols for these same vowels are, respectively, [ɨ], [ï], [ə], and [ë]. Individual descriptions of Cariban languages have generally not distinguished the central from back articulation (no individual language makes a phonemic distinction between them), so the symbols [ï] and [ë] have often been used to indicate central vowels. Meira's (2002) usage is this latter; our own usage of [ï] and [ë] in section 4-5 reflects the second, the Americanist tradition.

- Across the family, *i-ô > *e*
- In all branches of the family except Parukotoan, *jô > *je*
- In a few languages, *ô > *a* /__C*a* (vowel harmony).
- In a few languages, *ô > *a* /#__ (initial lowering, largely restricted to high-frequency morphemes like *ô- '2', *ôte- 'detransitivizer', and free pronouns).
- There may also be some idiosyncratic resolutions to vowel hiatus when *ô is involved (*e-ô > *e*, but with any other vowel, *e-V > V; also, in some cases, *ôe > *ô*).
- In a later development, in about half the languages of the family, *ô > *o*, creating *e ~ o* ablaut from *e ~ ô* ablaut.

At this point, we arrive at the contributions of this paper: we collect cognate sets from all members of the family for which we have data (Appendices 2 and 3) and we consider (section 3) all the modern reflexes of *ô that appear in the resulting correspondence sets (Appendix 1). A casual inspection of these cognate sets makes it clear that we have a distinct, reliable correspondence that must be reconstructed as a separate proto-phoneme. We then construct the case for evidence of an attested change: *ô > *ë > *o* in Kari'nja (section 4). We conclude with some discussion of the probable phonetic value of Proto-Cariban *ô (section 5).

3. MODERN REFLEXES OF *ô. The correspondences are presented in Appendix 1, extracted from the cognate sets in Appendix 2. We divide these into two groups of languages: those in which the modern reflex of *ô represents a separate phoneme (section 3.1 below) and those in which the modern reflex of *ô has undergone an unconditioned merger with another phoneme (section 3.2). We then consider briefly some individual phonological environments in which a number of modern reflexes of *ô have undergone conditioned mergers with other phonemes (section 3.3).

As a preliminary to the consideration of individual language names, we present a conservative modern classification of the family (Figure 1). We have chosen specific spellings for the name of each language based on (i) the spelling preferred by the majority of the communities who speak the language (where that is known to us) and (ii) the spelling adopted in the source(s) we cite. To help avoid adding to confusion amidst the proliferation of names, we specify some commonly-known alternate names for certain languages: Karihona (Carijona), Kari'nja (Carib, Galibi, Cariña), Katxuyana (Kaxuyana), Kuhikuru (Kuikuro), Kapóng (preferred spelling in Guyana, where most Kapóng live, for Akawaio, Patamuna and Ingarikó), Pemón (preferred spelling in Venezuela, where most Pemón live, for Taurepang, Arekuna and Kamarakoto), Yabarana (for Yawarana, Mapoyo/Wanai, and Pémono), Tiriyó (preferred spelling in Brazil, as opposed to Trio, preferred spelling in Suriname), and Ye'kwana (De'kwana, Makiritare, Maionggong). Other names are more standard, at least in the modern literature.

We believe that our classification is reliable in recognizing individual languages and the immediate Groups into which many of them fall. Among the Groups, only Taranoan has

been thoroughly documented (Meira 2000); the others are so obviously related that we do not expect additional documentation and comparative work to alter them.

Relationships above the level of Group are more tentative, in part because of the relatively conservative nature of Cariban phonology (which led Girard (1971) not to posit any higher-level branches) and in part because we have not yet thoroughly combed our new databases in search of potential shared innovations. The Venezuelan Branch presented here differs from the one proposed in Gildea (2003) and Mattéi-Muller (2002, 2003) in that it excludes the Kumaná and Makiritare Groups, which were originally included due in part to the mistaken belief that their separate phoneme *ə* might constitute a shared innovation (from Proto-Cariban *o). While additional criteria might still relate them to the other Venezuelan Branch languages, here we strip the branch to those languages that present more potential shared innovations. The Pekodian Branch is as proposed in Meira & Franchetto (2005). Certainly future comparative research is likely to identify more high-level relationships amongst the Groups.

FIGURE 1. A conservative classification of the modern Cariban family

Venezuelan Branch (A-B-C-D)

Pemóng-Panare Macro-Group (A-B)

A. Pemóng Group (Kapóng [Akawaio, Patamuna, Ingarikó], Makushi, Pemón [Taurepang, Kamarakóto, Arekuna]).

B. Panare

Mapoyo-Tamanaku Macro-Group (C-D)

C. Yabarana (Mapoyo, Wanai, Yawarana, Pémono)

D. †Tamanaku

Pekodian Branch (E-F)

E. Bakairí

F. Arara Group: Arara (Parirí), Ikpéng (Txikão)

Residue (Groups and Languages still in search of branches, in alphabetical order)

Groups

G. †Kumaná (†Chaima, †Cumanagota)

H. Ye'kwana (De'kwana, Makiritari, Maiongong)

I. Nahukwa Group: Kuhikuru (Kuikúru, Kuikuro), Kalapalo

J. Parukotoan Group

J1. Katxúyana (Kaxuyana, Shikuyana, Warikyana)

J2. Waiwai SubGroup: Waiwai (Wabui, Tunayana), Hixkaryana

K. Taranoan Group

K1. Tiriyó Subgroup: Akuriyó, Tiriyó (Trio)

K2. Karihona

L. Yukpa Group: Yukpa, Japréria

Languages

M. Apalaí

N. Kari'nja (Carib, Kalinya, Cariña, Galibi)

O. Waimirí Atroarí

P. Wayana

3.1. *ô AS A SEPARATE PHONEME. Eleven modern languages present clear cases in which modern reflexes of *ô are a separate phonemic category, ten presenting the mid-central vowel *ə* and one the mid front vowel *e*. In addition, Jesuit descriptions of Tamanaku and Kumaná are best interpreted as reflecting mistranscriptions of *ə* (cf. Mattéi-Muller and Henley's (1990) work on Tamanaku, and cf. the correspondences for *ô in Appendices 2 and 3), raising the total number of languages that present *ə* to twelve. As listed in Figure 2, these languages belong to nine different groups, plus one language (in a group of one) that presents *e*, meaning a separate phoneme is attested in ten out of sixteen groups of the family. If the Venezuelan and Pekodian Branches are counted as only one unit each, then a separate phoneme reflex of *ô is seen in seven of twelve units, making it nearly impossible to posit that modern *ə* is a consequence of a shared innovation in which a single proto-phoneme (e.g., *o) split to become two modern phonemes in these languages.

FIGURE 2. Languages with *ô as a separate phoneme

Mid-central vowel: *ô > *ə*

A. Pemón, Kapóng — Venezuelan Branch
B. Panare
C. Yabarana/Mapoyo
D. (Tamanaku)
E. Bakairi — Pekodian Branch
G. (Kumaná)
H. Ye'kwana
K. Tiriyó, Akuriyó, Karihona
P. Wayana

Front vowel: *ô > *e* (*e merges with *i)

I. Kuhikuru

3.2. UNCONDITIONED MERGERS. Ten modern languages present clear cases in which modern reflexes of *ô have consistently merged with another proto-vowel, and thereby do not constitute a separate phonemic category, eight presenting as the modern reflex the mid back rounded vowel *o* and two the high central unrounded vowel *ɨ*. As listed in Figure 3, these languages belong to seven different groups (five merging *ô with *o, two with *ɨ). Note that two members of the Pekodian Branch and one member of the Venezuelan Branch are present in this group as well, making it difficult to argue that either merger can be characterized as either entirely a shared innovation or as due to contact.

FIGURE 3. Languages in which *ô merged with another vowel

Merge with *o
- F. Arara, Ikpéng ——— Pekodian Branch
- J. Waiwai, Hixkaryana, Katxúyana
- L. Yukpa
- M. Apalaí
- N. Kari'nja

Merge with *ɨ
- A. Makushi ——— Venezuelan Branch
- O. Waimiri-Atroari

3.3. CONDITIONED MERGERS (MEIRA, GILDEA & HOFF FORTHCOMING). In our ongoing investigation of modern Cariban morphophonology, we have found a number of phonological processes that appear to apply only to modern reflexes of *ô. Particularly fronting leading to ablaut, lowering harmony, and word-initial lowering. The same fronting, harmony processes, and sporadic lowering appear to have operated historically to yield specific inconsistencies in modern cognate sets.

First, we recapitulate the findings of Meira, Gildea & Hoff (forthcoming) with regard to ablaut caused by fronting. In all languages of the family, vowel hiatus caused by *ɨ- '3' prefixed to any root beginning with *ô results in a single vowel, *e, at the beginning of the root; before all other vowels, *i- simply disappears with no synchronic reflex. In all except the languages of the Parukotoan Group (Hixkaryana, Katxuyana, Waiwai), the sequence *jô > *je*, most frequently following the relational prefix *j-*, but also morpheme-internally in 3 lexical items (cf. 'tooth', 'bone', 'cook' in Appendix 2). In no Cariban languages does *i- '3' survive preceding vowels, and in most modern Cariban languages, word-initial *j- 'Rel' > Ø, which leads to the loss of the conditioning environment for fronting. As such, roots beginning with the vowel *ô present a synchronic alternation that Meira, Gildea & Hoff call *ablaut* in these languages (Table 2).

Table 2. Ablaut with the word for 'eye' in selected Cariban languages (back grade shaded)

	P.C. *ônu	Tiriyó *enu*	Kari'nja *enu*	Wayana *ewu*	Bakairi *enu*	Katxuyana *enu*
3	*i-ônu	*enu*	*enu:-ru*	*ewu*	*enu*	*enu-ru*
2	*ô-j-ônu	*ə-enu*	*aj-e:nu-ru*	*əw-ewu*	*inu*	*o-onu-ru*
1	*u-j-ônu	*j-enu*	*j-e:nu-ru*	*j-ewu*	*j-enu*	*j-onu-ru*
NP	*j-ônu	*enu*	*enu:-ru*	*ewu*	*enu*	*j-onu-ru*
1+2	*k-ônu	*k-ənu*	*k-onu:-ru*	*k-əwu*	*k-ənu*	*k-onu-ru*
3R	*t-ônu	*t-ənu*	*t-onu:-ru*	*t-əwu*	*t-ənu*	*t-onu-ru*
Ø	*ônu	*ənu*	*o:nu*	*əwu*	——	——

The rows are sorted according to ablaut effect. The top row shows that the initial *ô in 'eye' is fronted to *e* in all languages for third person. The next three rows demonstrate the differential effect of *j- 'REL', which only occurs following the first and second person prefixes and following a preceding possessor NP: *ô is fronted to *e* for most languages, excepting Katxuyana (a Parukotoan language, in the final column), where *ô is not fronted following *j-, but remains back and later participates in the general merger when *ô > *o*. The bottom three rows demonstrate the form of the root taken when there is neither a third person nor a relational prefix conditioning fronting: the unfronted reflexes of *ô are simply unrounded *ə* or rounded *o*.[5]

The second conditioned merger is an outcome of lowering harmony triggered by the low central vowel *a* as the nucleus of the following syllable: *ô > *a* / __C*a*. This is described in multiple languages as a synchronic vowel harmony rule, in which a prefix contains a modern reflex of *ô (whether *ə* or *o*) that becomes *a* when the prefix is attached to a root whose first vowel is *a* (both with and without an intervening consonant). Ready examples are found in languages where the second person prefix has the allophone *a-* conditioned by an *a* in the initial syllable of the root (e.g., Apalai, Koehn & Koehn 1986:97; Hixkaryana, Derbyshire 1985:199; Katxuyana, Gildea's field notes; Tiriyó, Meira 1999:201), or where the detransitivizer prefix has an allomorph *at-* or *as-* when the root has *a* in the first syllable (Apalaí, Arara, Ikpéng, Bakairi, Hixkaryana, Karihona, Katxuyana, Kuhikuru, Makushi, and Ye'kwana; Meira, Gildea, & Hoff forthcoming). The most extreme example in this case is found in Katxuyana, where vowel harmony applies across the entire stem (i.e., no stems are attested that begin *o*C*a*) and where some cases of ablaut are between fronted *e*

[5] Recall that in Kuhikuru, a later vowel shift led to *e > *i* and *ô > *e*, such that the ablaut alternation is not *o* ~ *e*, but rather *e* ~ *i*: e.g., *egi* 'song', *ete* 'village', *eku* 'semen' → *u-igi-sɨ* 'my song', *u-itu* 'my village', *u-iku-ɣu* 'my semen'.

alternating with lowered *a*, for example, *e̲wahu* 'his calf (of leg)' (< *i-ôwapu) versus *t̲-awahu* 'his own calf' (< *t-ôwapu).

In addition to harmonic lowering, there are some inconsistent cases of *ô > *a* word-initially (often a prosodically weak position in these dominantly iambic languages). These are gathered together in Appendix 3. We will return to this phenomenon in more detail in section 4, where we examine both modern dialectal variation and attested historical transcriptions of Kari'nja.

3.4. SUMMARIZING THE EVIDENCE SO FAR. At this point, we can see two converging lines of evidence pointing to Proto-Cariban *ô as a distinct vowel: first, there is a substantial set of cognates that display a consistent correspondence of *ə* : *o* : *ɨ* : *e*, and second, there are two conditioned historical phonological processes (fronting and lowering) that have affected only this vowel. We can now turn to these changes and try to sequence them, after which we can look for evidence of the relative time depth of each change.

The most obvious conclusion available to us is that unconditioned mergers all came after any conditioned changes, as subsequent conditioned changes would have applied to the entire merged category, and not just the subset that has its origin in *ô. Hence, we date the many cases of unconditioned *ô > *o* and the two cases of unconditioned *ô > *ɨ* as the last changes to take place. The order of the conditioned changes is less transparent, but it seems clear that fronting is an exceptionally old process: coalescence of the third-person prefix *i- with the initial vowel of the stem probably predates Proto-Cariban, and the ubiquity of the change *jô > *je* argues for its antiquity as well, especially given that the conditioning environment (the initial *j-) is lost in so many languages. In contrast, both harmonic and sporadic lowering of *ô > *a* is more limited, either to specific very old, high-frequency morphemes such as *ô- '2' and *ôte- 'DETRANSITIVIZER [RECIPROCAL]', or limited to a few languages in a geographically contiguous area of northern Brazil.[6] Of particular interest is the evidence from Katxuyana that fronting predates lowering; only after *i-ô > *e* for third person do the remainder of *ô lower to *a* in the harmonic environment, creating *e* ~ *a* ablaut for a handful of words.

Given that the mergers are at the end of this chain of changes, one might ask how recently such mergers have taken place. The answer is "relatively recently", as seen in the divergence in outcome within closely-related genetic groups. For example, Makushi is still largely mutually intelligible with its two sister languages in the Pemón group, Pemón and Kapóng, yet of the three, only Makushi has merged *ô > *ɨ*. Meira & Franchetto (2005) argue for the existence of the Pekodian Branch of the family, uniting Bakairi with the Arara/Ikpéng Group, yet while Bakairi maintains *ô as the separate phoneme *ə*, Arara and Ikpéng have both merged *ô > *o*. However, it is difficult to project the time depth of these changes back in real time, as we cannot determine actual dates when the groups in question separated. In section 4, we turn to something closer to a case of attested change: the varieties of Kari'nja captured in historical wordlists and attested in modern dialectal variation.

---

[6] For now, we offer this maximally restrictive characterization, but as more detailed descriptions become available for other languages, productive harmonic lowering may be found outside of this region.

4. THE KARI'NJA EVIDENCE: *ô > *o* AS A RECENT CHANGE. The Kari'nja language, also known as Carib proper, Galibi, Kaliña, Kari'na, and Cariña, is the one for which the family received its name. It is spoken by between 10,000 and 25,000 people along a roughly 1000-mile arc of the coast and one to two days' travel inland beginning in Brazil near the border with French Guiana, continuing through French Guiana, Surinam and Guyana, and ending in the easternmost 300 miles of the Venezuelan coast. As one of the first languages contacted, and as a language spoken by a larger population than most, Kari'nja has a rich record of colonial documentation, plus more ample modern documentation from multiple dialects. Present-day Kari'nja is one of the languages in which *ô merged with *o, generating a typical six-vowel system (*a e i o u ï*). There is, however, some historical evidence that points to the existence of an earlier seventh vowel in the Kari'nja vowel system.

Further, Kari'nja served as the basis of a pre-Colombian trade pidgin that is documented in two word lists. When the West-European traders arrived on the north coast of South America, they found a pidginized Kari'nja language waiting for them to be used in their dealings with the Indians, all along the coast (Boyer 1654, Biet 1664). And long before that time, due to a Kari'nja invasion some two to three centuries before the arrival of Colombus, the same contact language had already served as the lexifier for the male speech of Island Carib, an Arawakan language of the Lesser Antilles with a strong gender-based register distinction (Taylor & Hoff 1980, Hoff 1995). This language is known to us thanks to the work of Raymond Breton (1665, new edition 1999). A modern descendent of Island Carib, Garifuna, received some descriptive attention in the 20th century (Taylor 1951, 1977).

Taken together, these sources give evidence that early Kari'nja retained a phonemic distinction between the reflex of *ô as an unrounded mid back vowel *ë* and that of *o as *o*. In all modern dialects of Kari'nja, *o is still *o*, but an independent reflex of *ô has completely vanished: most of its territory has been taken over by *o*, a small part by *a*, and via the ablaut process, another part to *e*. We consider first the evidence provided by Island Carib for the survival of *ô as an independent phoneme in early Karinya (section 4.1). Then we turn to the process by which some instances of *ô merged with the phoneme /a/ in the Kari'nja pidgin and in some dialects of modern Kari'nja (section 4.2).

4.1. ISLAND CARIB AND GARÍFUNA: THE EVIDENCE FOR EARLY KARI'NJA *ë < *ô. All lexical information about Island Carib comes from Breton's (1665) dictionary. It is not a trivial task to interpret Breton's orthography, in which one finds several symbols that suggest an *e*-like sound—*e, eu, ê, é*—and also the symbols *ou* and *o*, which suggest back rounded vowels. Breton's data have been successfully decoded by Douglas Taylor, largely on the basis of his findings from descriptive study of present-day Island Carib, better known as Garifuna (Taylor 1951:12, 160-171; 1977: 29-43, 138-142). We illustrate Breton's transcription and Taylor's interpretation of it by means of the concrete examples found in Table 3.

In Table 3, cognates are presented from modern Kari'nja, Island Carib, and modern Garifuna, and from these we reconstruct Early Kari'nja, the common ancestor that is shared by modern Kari'nja and, via Island Carib, by modern Garifuna.

Table 3. Reflexes of *ô,*ï, *e, and *o in early and modern stages of Kari'nja[7]

		Modern Kari'nja	Early Kari'nja	Island Carib		Garifuna
1	'two'	/oko/	*/ëkë/	*eukê*	/ïkï/	-----
2	'stone'	/topu/	*/tëpu/	*tébou*	/tïbu/	/dïbu/
3	'flea'	/siko/	*/sikë/	*chicke, -chigu-*	/sikï/	/sigï/
4	'mountain'	/wïpï/	*/wïpï/	*ouébo*	/uïbu/	/uïbu/
5	'axe'	/wïwï/	*/wïwï/	*houéhoue*	/uïuï/	-----
6	'sun'	/weyu/	*/weyu/	*huéyou*	/ueiu/	/ueiu/
7	'person'	/itoto/[I]	*/itoto/	*etoútou*[II]	/itutu/	/idudu/[III]
						[I]Tiriyo Indian, [II]Arawak Indian, [III]Miskito Indian.

Starting from Early Kari'nja *ëkë* 'two' in the third column and moving to modern Kari'nja in the column to the left, we see the result of the historical rounding process in both vowels. Moving to Island Carib in the column on the right, Breton's version *eukê* together with the comments in his *grammaire* (1667) suggest an *e*-like rather than an *o*-like sound. Taylor came to the conclusion that one and the same phoneme of Island Carib is symbolized here once by *eu* and once by *ê*, and that this phoneme was back and unrounded. It cannot have been high or mid in a phonological sense, because the high/mid distinction is not available to back vowel phonemes in Arawakan systems like that of Island Carib: rounded /u/ presents the allophones [u, o], whereas the actual pronunciation of unrounded /ï/ varied between [ï, ë]. Breton distinguished the high from mid allophones, writing both *o* [o] and *ou* [u], *ê* [ë] and *eu* [ï]. Taylor, observing the allophony in the modern language, recombines these distinct graphemes to the single phoneme they must have represented.

In the second and third rows, Modern Kari'nja shows the expected rounded reflexes. However, the spelling *tébou* for 'stone' reveals a typographically vulnerable spot in Breton's book: a circumflex has been suppressed here by the stress mark, creating typographical ambiguity between the front and back unrounded mid vowels–this happened quite often in the dictionary, and can only be disambiguated by consulting Taylor's modern Garifuna examples. In the third example, Breton's unmarked *e* suggests a front vowel, but both the modern Kari'nja reflex and the modern Garifuna reflex suggest that it must have been the unrounded back vowel. As noted in section 2, such confusion is common in the Jesuit sources; even though Breton clearly could hear (and did mark) many /ë/ and /ï/, his data are not free of this problem.

[7] Note that we deviate from one of Taylor's spelling conventions: for the non-low, unrounded back vowel phoneme of Garifuna we write /ï/ instead of Taylor's /o/.

In 'mountain' (row 4) and 'axe' (row 5) we assume that a high unrounded back /ï/ of early Kari'nja remained unmodified in modern Kari'nja and Island Carib (and in Garifuna).

Row 6, 'sun', in Island Carib contained a front vowel /e/, again made ambiguous by the stress mark. In comparing rows 5 and 6, we can see that Breton leaves us a clue to vowel quality in the spelling of the preceding non-syllabic /u/. Like his compatriots Boyer and Biet, Breton transferred a peculiarity of French phonology and French orthography to his Kari'nja data. French distinguishes between a labial-palatal approximant (as in /ɥi/ *huit* 'eight') and a labial-velar approximant (as in /wi/ *oui* 'yes'). Accordingly, the French authors spell the labial approximant of Carib before a front vowel as *ü, hü* or *vü*, and before a back vowel as *oü*. So we can be certain that *houé* (in row 5) indicates the sequence /uï/, whereas *hué* (as in row 6) indicates the sequence /ue/, both confirmed by the cognates in modern Kari'nja and Garifuna.

Row 7 illustrates reflexes of *o: they remain /o/ in modern Kari'nja, but in both Island Carib and Garifuna, they have merged with /u/ into a single phoneme, which (following Taylor) we represent as /u/.

To summarize the Island Carib facts, a phonemic contrast is seen between two back vowels, one rounded and one unrounded, each of which apparently varied freely between high and mid pronunciations. In Breton's Island Carib transcriptions, the back unrounded vowel is sometimes uniquely distinguished as *ê* or *eu*, but often ambiguously written as *é* or *e*; these can sometimes be disambiguated via Breton's convention for writing the bilabial approximant, rendering the sequence /uï/ as (*h*)*oué* and the sequence /ue/ as *hué*.

We now summarize the argument for our reconstruction of four back vowels in Early Kari'nja. First, modern Kari'nja presents three back vowels, two rounded /o, u/ and one unrounded /ï/. Second, the Island Carib rounded back vowel /u/ contains correspondences to both modern Kari'nja /u/ and a subset of /o/. We presume two Early Kari'nja rounded back vowels, high *u and mid *o, which survived unchanged into modern Kari'nja, but which, upon being borrowed into Island Carib's Arawakan phonological system, merged into a single back rounded phoneme. Third, the Island Carib unrounded back vowel /ï/ contains correspondences to both modern Kari'nja /ï/ and another subset of /o/. We presume two Early Kari'nja unrounded back vowels, high *ï and mid *ë, which underwent an analogous merger in Island Carib, to a single back unrounded vowel; in modern Kari'nja, the high back unrounded vowel survived unchanged, but the mid back unrounded vowel /ë/ became rounded in all environments and thereby underwent an unconditional merger with *o.

Finally, we have to link up early Kari'nja *ë to proto-Cariban *ô. To test this connection, we added as many Island Carib and Garifuna cognates as possible to our Cariban cognate sets in Appendix 2. The resulting correspondences, presented in Appendix 1, show that, although we were unable to find clear evidence for the actual Island Carib vowel in most cognates, in the handful of cases where we have such evidence (whether from Taylor's decoding of Breton's Island Carib orthography or from Taylor's modern Garifuna data) the reflex of Proto-Cariban *ô is the unrounded back vowel. Thus, the outcome of this experiment confirmed our expectations: where we have evidence, early Kari'nja *ë is the continuation of Proto-Cariban *ô. We summarize these changes in Table 4.

Table 4. Proto-Cariban (PC) > Early Kari'nja (EK) > attested languages

PC	EK	Kari'nja	Island Carib	Garifuna
*u	*u	*u*	*u* [u~o]	*u* [u~o]
*o	*o	*o*	*u* [u~o]	*u* [u~o]
*ɨ	*ï	*ï*	*ï* [ï~ë]	*ï* [ï~ë]
*ô	*ë	*o*	*ï* [ï~ë]	*ï* [ï~ë]

4.2. *ô > *a* BEFORE AND SINCE EARLY KARI'NJA. In the opening sentence of section 3.2 (and again in the opening paragraph of section 4), we presented the somewhat simplistic categorical statement that the loss of *ô in present-day Kari'nja was due to its merger with *o*. However, prior to this generalized merger was another, more restricted merger, in which a subset of Proto-Cariban *ô > *a*. In section 4.2.1 we describe a morphonological alternation of /a/ and /o/ in the Tïre'wuyu dialect of Maroni and Mana Rivers that synchronically continues to manifest *both* shifts. In section 4.2.2 we present lexical data from the old printed sources, which suggest a possible shift from *ë > *a* > *o*. And prior to both of these shifts was the shift mentioned briefly at the end of section 3.3, in which both Island Carib and Kari'nja share a number of words in which *a* is the reflex of Proto-Cariban *ô in initial position. Given that all modern Kari'nja dialects and the borrowed lexicon in Island Carib share the same reflex, we presume that for these words, the shift from *ô > *a* must have happened between Proto-Cariban and Early Kari'nja (cf. Appendix 3).

4.2.1. MORPHONOLOGICAL ALTERNATIONS IN THE TÏRE'WUYU DIALECT OF KARI'NJA: *a ~ e, o ~ e, a ~ o*. The Tïre'wuyu dialect differs from the other Kari'nja dialects in Suriname by the occurrence of /a/ (instead of /o/) in prosodically weak first syllables.[8] Where syllables are prosodically strong, /o/ occurs in all dialects, and ablaut changes *ô > *e* the same in all dialects. The paradigms in (1-4) illustrate, (1) and (2) with nominal, (3) and (4) with verbal examples.

[8] Diphthongs and vowels followed by a consonant coda count as strong. Another source of strength is a general prosodic rule, which imposes either an iambic or a trochaic pattern on the first two syllables of words, starting from the left. In the western dialects, in the first foot this prosodic strength is phonetically realized only by quantity. The Tïre'wuyu dialect, however, also differs in *this* respect. In the first foot, too, strong prominence is primarily due to melodic rises or falls (Hoff 2000, unpublished ms). In the following examples, we shall indicate prosodically strong vowels by bold font.

(1)	**Tïre'wuyu**	**Other dialects**	**gloss**
	*ar**e**mi*	*oremi*	'spirit song'
	*k-a**r**e**e**mi-rï*	*k-o**r**emi-rï*	'our spirit songs'
	*y-er**e**mi-rï*	same	'my spirit song'
	*Ø-er**e**mi-rï*	same	'his spirit song'
(2)	***om**bata*	same	'face'
	*k-**om**bata-rï*	same	'our faces'
	*y-**em**ba-tarï*	same	'my face'
	*Ø-**em**bata-rï*	same	'his face'
(3)	*am**a**-no*	*oma-no*	'way of life'
	*k-am**ai**-ya*	*k-omai-ya*	'we live, dwell'
	*y-**e**mai-ya*	same	'I live, dwell'
	*Ø-em**a**mï-rï*	same	'his living, dwelling'
(4)	***om**bakano*	same	'waking up, transitive'
	*k-**om**bakae*	same	'you wake me up'
	*y-**em**bakano*	same	'he wakes me up'
	*kïn-**em**bakano*	same	'he wakes him up'

In all four examples, most dialects show the standard outcome of the old *o* ~ *e* ablaut process; however, in Tïre'wuyu, only the vowels in heavy syllables (2, 4) show the *o* ~ *e* alternation, whereas the vowels in light syllables (1, 3) instead show an *a* ~ *e* alternation. Both pairs, of course, continue the same historical ablaut process, but the Tïre'wuyu dialects show a prior change of unstressed *ë > *a* before the unconditioned merger of *ë > *o*.

To witness an alternation *a* ~ *o*, we must exclude ablaut by avoiding fronting prefixes. To this end, we choose one verb from an intransitive subset that employs only non-fronting prefix allophones. It is a derivation by means of the intransitivizer prefix, reconstructed to Proto-Cariban as *ôte-, with reflexes of *at-* in Island Carib, Tïre'wuyu and Venezuelan Kari'nja (Mosonyi 1978:61), but as (*w*)*ot-* in the other dialects of Suriname (Hoff 1968:122-4). See Table 5.

Table 5. The intransitivizer prefix at-/(w)ot-

	Island Carib	Tïre'wuyu dialect	Other dialects
'he propelled himself'	*n-a̲t-alima-i*	*n-a̲t-arima-i*	*n-o̲t-arima-i*

The paradigm in (5) shows that the vowel in the prefix remains *a* as long as it is prosodically weak, by being in the first syllable of the first, iambic, foot. When further prefixation (in the fourth example) pushes it into the strong position, *a* is replaced by *o*. Note that in the other dialects /a/ does not occur at all in the same prefix.

(5)	**Tïre'wuyu**	**Other dialects**	**gloss**
	*a̲t-**a**rima*	*wo̲t-arima*	'propelling oneself'
	*w-a̲t-**a**rima-i*	*Ø-wo̲t-arima-i*	'I propelled myself'[9]
	*n-a̲t-**a**rima-i*	*n-o̲t-arima-i*	'he propelled himself'
	*kï-n-**o̲**t-arimano*	same	'he propels himself'

4.2.2. THE PATH FROM *ô TO *a* IN THE LEXICON. As seen in Appendix 3, several languages present *a* as the reflex of Proto-Carib *ô when it occurs in the (unstressed) first syllable. The agreement of Island Carib, Garifuna, and all dialects of modern Kari'nja allow us to posit that Early Kari'nja was one such language. However, in considering more closely the data from Breton's dictionary with the equivalents in modern Kari'nja of Surinam, we found other cases in which an *a* of Island Carib coincided with an *o* in modern Kari'nja, as well as other anomalous correspondences that might represent modern reflexes of Proto-Cariban *ô. This section is dedicated to an exploration of these anomalous correspondences.

Presumably, changes towards /a/ have left few traces in the present language. From the written sources, however, one may gain the impression that in the past the frequency of *a* from *ô has been higher. We believe that this impression in most, though not all, cases would be incorrect. With the intention to gain more insight into what historically may have happened, we bring together a number of words that in the modern language of Surinam all contain one or more *o*'s (cf. the last column in Tables 6-7, next-to-last column in Table 8), corresponding to what one or two old sources wrote as either *e* or *a*.

Tables 6-8 each present data from Island Carib, Carib Pidgin, Modern Kari'nja as spoken in 1655, and Modern Kari'nja as spoken in 1955. One major divide separates the first two columns (both reflecting a contact language that used Early Kari'nja as a lexifier) from the third and fourth columns (representing different historical stages of direct descendents of Early Kari'nja). A second major divide separates the first three columns (all 17th-century sources of questionable reliability) from the fourth (a reliable modern source). Within the Kari'nja language itself, Pelleprat's (1655) materials indicate that the merger of *ô > *o* was already nearing its completion. But the two contact languages on the left, though recorded at nearly the same time, reflect the forms of the words as they were many centuries earlier, when the Pidgin took its lexical material from Early Kari'nja and thereby was cut loose from further developments within that language (Boyer 1654, Biet 1664). Basing himself on archeological evidence, Boomert (personal communication) estimates the beginning of widespread intertribal trade in the Caribbean area, and with it the formation of the Pidgin,

[9] Suffixation with *-i* expresses involvement mode, *-no*, in the next example, the unmarked (present) tense.

at a period between 500 and 800 AD.[10] The divide between the Pidgin and its lexifier language became still steeper when the latter was incorporated into the Arawakan Island Carib language, perhaps two to three centuries before the arrival of Colombus.

In the three columns that are based on the old written sources the original spellings have been retained. Only Breton made an attempt to distinguish the unrounded back vowel /ë/ from /e/ (as discussed in section 4.1) by rendering it either as a digraph *eu* (items 9 and 11) or as *ê* (items 1, 2, and 10). Where these unique graphemes are lacking, we still identify *e* or *é* as /ë/ when it follows *oü* (item 5) or when it takes part in the historical developments towards *a* and *o* (the remaining underlined *e* or *é* in 3, 4, 6, cf. also Appendices 2 and 3). Additionally, we see a few cases in which Breton appears to represent /ë/ via digraphs that would have yielded a centralized nasal vowel in French: *an* (6) and *um* (7). However, even after resolving these orthographic convolutions, we remain with a residue of cases (8-13), in which Breton transcribed *a* for what the comparative record (and the outcome in Modern Kari'nja) tells us should have been the phoneme /ë/.

Looking now at the two middle columns, we can see that Boyer and Biet's Pidgin data show, if anything, more cases of *a* than Breton's Island Carib, whereas Pelleprat's contemporary Kari'nja words show the same *o* as Hoff's modern data.

[10] Boomert relates the beginning of intertribal trading to the Koriabo pottery complex that appears in the Guianas at an estimated date between 500 and 800 AD. "Koriabo pottery is found all along the coast of the Guiana's, and it is also richly represented in the interior. Shards of traded pottery have been found even in the lower reaches of Orinoco R. In Brazilian Guiana and on Oyapock R. a different but closely related pottery complex is found. Therefore we may safely surmise that this whole area was in constant interaction, in any case along the coast. This also applies to the relations between the coastal area of the Guiana's and the Windward Islands. As a contact language presupposes relations between Indians of various linguistic backgrounds, I believe that a contact language based on Carib should date from the late pre-historic period, say from between 500-800 AD" (p.c.; see also Boomert 1995).

Table 6. Clear reflexes of *ô in early and modern stages of Kari'nja

		Island Carib, 1655	Pidgin lists, 1654-1664	Modern Kari'nja, 1655	Modern Kari'nja, 1955
1	'pig (peccary)'	*boinkê*	*poingé, poinga, poingo*	*poinco*	*poingo*
2	'fart'	*i-bikê-li*	*piqua*	—	*piko*
3	'flea'	*chicke*	*(Dutch sika)*	chíco	siko
4	'stone'	*tebou*	—	*tóbou*	*topu*
5	'woman'	*oüélle*	*oüali*	*oüori*	*worïi*
6	'you'	*amánle*	*amoré*	*amóro*	*amoro*
7	'moon'	*nónum*	*nouna*	*noûno*	*nuno*
8	'who'	*anaki*	*anakè, anac, nec*	*nóke, anóke*	*noki*
9	'come!'	*akeu*	*ac-né*	*occó-né*	*ohko ne*
10	'how many'	*átêli*	—	*óttoro*	*ohtoro*
11	'cut'	*ch-ackeuta-é*	—	—	*s-akoto-ya*
12	'you came'	*mábouica*			mopïi ko
13	'sleep'	*aónikay*	*nanegué*	—	no'nïkïi
14	'star'	*chiric*	*serica, sirica*	*sirícco*	*siriko*
15	'he'	—	*inali*	—	*inoro*
16	'snake'	—	*acoïou*	—	okoyu
17	'toad'	—	*balalou*	—	*pïroru*

Table 7 offers another nine potential modern reflexes of *ô—wherever a Kari'nja /o/ corresponds to an *e*, *a*, etc. in Island Carib we add them to this table, even though there is no evidence for the status of the vowel in the larger comparative Cariban picture. Again, we find Island Carib *e* (1-3), and *a* (4-6) corresponding to Kari'nja *o*, and once again, the Pidgin list shows nearly all *a* for these correspondences; here, Pelleprat's (1655) Modern Kari'nja data are missing most cognates, with only one case of *o* (3) to match those in Table 6 and an anomalous case of *e* (1).

Table 7. Possible reflexes of *ô in early and modern stages of Kari'nja

		Island Carib, 1655	Pidgin lists, 1654-1664	Modern Kari'nja, 1655	Modern Kari'nja, 1955
1	'angry'	*te̲re-coù*	*ta̲ri-qué, te̲ri-qué*	*té̲le-ké*	*to̲reh-ke*
2	'family'	*e̲me-ri*	*ï-a̲mo-ri*	—	*o̲morï*
3	'wind'	*bebéite̲*	*epepeita̲*	*bebeíto̲*	*pepeito̲*
4	'shiver'	*tica̲main*	*tiga̲miné*	—	*tïko̲:mïine*
5	modal particle	*á̲la̲*	—	—	o̲ro̲
6	'water-spirit'	*a̲cáyouma̲n*	—	—	*o̲ko̲:yumo̲*
7	'day after tomorrow'	—	*(a)ma̲nicoropo*	—	*mo̲nïngo-ropo*
8	'thievish'	—	*ma̲namé*	—	*mo̲name*
9	'long'	*mouchí-pe*	*ma̲nci-pe, mo̲sim-bè*	—	*ma̲nsi-pe mo̲:si:-pe*

Finally, we have three anomalous cases, presented in Table 8. Two clear reflexes of Proto-Cariban *o present unrounded reflexes in Island Carib: 'beer' presents the sequence *oue* /uë/, which is confirmed to be the unrounded back vowel by its Garifuna cognate *uïgu*, and which also presents a correspondence of *a* in the Pidgin; 'cashew' presents the expected reflex *ou* /u/, but alongside it also the unexpected *á*. 'Bullet' presents a still more interesting story, in which the final vowel of the Spanish loanword *pelota* is borrowed into both the Pidgin and Modern Kari'nja, as *e* (presumably /ë/) in the Pidgin, but as *o* in Modern Kari'nja.

Table 8. Parallel correspondences that are not reflexes of *ô

		Island Carib, 1655	Pidgin lists, 1654-1664	Modern Kari'nja, 1655	Modern Kari'nja, 1955	PC
1	'beer'	*oüe̲cou*	*oüa̲cou*	*oú̲o̲cou*	*wo̲ku*	*wo̲ku
2	'cashew'	*o̲ú̲loüi, á̲loi*	—	—	*o̲roi*	*o̲roi
3	'bullet'		piroté	piróto̲	pïroto̲	pelota̲ (Sp)

How do we account for the patterns that involve *a* in Island Carib and in the Pidgin? First, are we forced to accept another wave of unconditioned sound change, in which /ë/ > /a/? Second, especially given the case of Spanish /a/ being borrowed into Modern Kari'nja as /o/, is it possible that /a/ might have served as a stepping-stone in the change from /ë/ to /o/?

Turning to the first question, we must consider whether we accept Breton's, Boyer's, and Biet's transcriptions as accurate. For Boyer and Biet, it seems clear that they did not hear the vowel /ë/, possibly because the distinction did not survive so late into the Pidgin, but more likely because they (like the Jesuits working on Tamanaku and Kumaná) simply mistranscribed /ë/ as *a* or *e*. In support of the mistranscription position, note that the final vowel of *pelota* would almost surely have been borrowed into the Pidgin as either /a/ or /ë/ (the two available back unrounded non-high vowels); the *é* that is recorded could not represent a front vowel, and so must represent /ë/. Since Breton did have graphemes dedicated to /ë/, obviously he could hear the distinction at least sometimes; yet his use of *an, um,* and (unmarked) *e* suggests that he sometimes confused the distinction, and it is not unreasonable to assert that some cases of *a* might be similarly confused renditions of /ë/. On the other hand, Breton lived for many years among the Island Carib and spoke their language. What is more, in the two cases where Taylor provides a modern Garifuna form, the *a* in Breton is confirmed: item 6 from Table 7, *a̲cá̲youman* 'water-spirit', is confirmed by modern Garifuna *a̲ga̲iumau* 'water spirit', and item 12 from Table 6, *má̲bouica̲* 'you came', is confirmed by Garifuna *má̲buiga̲* 'so you have come'. We are forced to conclude that at least some portion—and perhaps all—of these anomalous *a*'s in Breton are the outcome of an unconditioned sound change, /ë/ > /a/.

However, even accepting the existence of /ë/ > /a/, we still find it difficult to believe that an /a/ in Island Carib or Pidgin could represent a transitional stage between /ë/ and /o/. In terms of articulation, even if we assume that lowering of /ë/ produced a temporary second low vowel phoneme, back /ɑ/ (distinct from front /a/), which could then change to /o/, this is quite a labor-intensive way to get from point a to point b, when simply adding lip rounding to /ë/ could do the job (cf. section 5). We prefer to assume that Frenchmen dealing with the Pidgin speakers identified the /ë/ of their Amerindian counterparts with the low back unrounded phoneme /ɑ/ of French (their own language), and accordingly used the grapheme a for both /a/ and /ë/.11 If the Amerindians also shared the assumed equivalence ('ë of the Amerindians = ɑ of the Europeans'), then we have an explanation for what happened to borrowed Spanish 'bullet': it must have independently started both its Pidgin and its Kari'nja careers as perotë (written piroté by Biet); in Kari'nja, it must have been among the last of the Kari'nja lexicon to contain a /ë/, which soon followed the rest of the phonemic category (the reflexes of *ô) in becoming /o/. Mosonyi has a few more entries with /a/ where Suriname Kari'nja has /o/: akoðu 'snake' (see acoïou, okoyu in row 16 of Table 6); amɨða 'beauty', for Suriname Kari'nja omïya. Presumably, the variants are the outcomes of independent shifts, which happened after the geographical separation of the Venezuelan and Surinamese dialects.

[11] Dutch and Spanish speakers may have done the same.

The cognates in Tables 6 and 7 also appear to give evidence for different stages of the wave in which *ë > *o*. Already in the very early period when the Pidgin was collecting vocabulary from Kari'nja, the wave must have started. For example, in the Pidgin most reflexes of *ô are *e* or *a*, but some *o* can also be seen: the second vowel of 'you' (row 6 of Table 6) and one variant of the final vowel of 'pig' (row 1 of Table 5); in Island Carib, the first root vowel of 'sleep' (row 13 in Table 6). In 'long' (row 9 of Table 7), *a* and *o* appear to have alternated through the centuries. *Mansipe* was noted down by Ahlbrinck (1931:271); the same form is also found for modern Venezuelan Kari'nja, see Mosonyi (1978:30).

We conclude with a final anomaly from Island Carib: the Proto-Cariban item *arô 'take, carry' appears in Island Carib with two variants, the expected *alee* (presumably = *alë*) and the surprising *eulê* (= *ëlë*). This latter form remains a puzzle.

5. TOWARDS DETERMINING THE THE PHONETIC VALUE OF PROTO-CARIBAN *ô. At this point, it should be obvious that Proto-Cariban *ô was an unrounded mid vowel—the only question is whether it was back /ë/ or central /ə/. In the first place, neither candidate can be excluded on general grounds. The mid unrounded back vowel is rare in the languages of the world and also peripheral typologically, but it is still found to exist in Scottish Gaelic and in Vietnamese (Ladefoged 2005:178-179; Ladefoged & Maddieson 1996:293). The central vowel is quite common, and while in some languages it is restricted to prosodically weak syllables, there are others where it functions on a par with the other vowels, for instance in Okanagan and in UK English (Jones 1960:88-91). In this section, we make the case for each of these alternants, beginning with /ə/.

In favor of /ə/, we put first the number of modern descriptions that have identified the modern reflex of *ô as a schwa. Although the practical orthographies generally utilize symbols like *ë* and *ö*, the articulatory descriptions that accompany them generally describe the vowel as central. Similarly, most descriptions of modern reflexes of the corresponding high vowel, ***ɨ**, describe it as central rather than back. Meira's acoustic analysis of the F1 and F2 of Tiriyó /ə/ and /ɨ/ confirm that they are both more central relative to the pair of back vowels, /u/ and /o/. To the extent that future acoustic studies confirm a parallel result in other languages with the seven-vowel system, the principle of parsimony would guide us towards reconstructing two central vowels. However, two-dimensional plots of formant frequencies alone are insufficient to determine the nature of vowel systems that combine the dimensions front/back and spread/round (Ladefoged 1971:74).[12] Once again, the comparative enterprise is dependent upon increasingly sophisticated fieldwork, including the instrumental study of articulation. Given that the community of fieldworkers in the Cariban language family continues to produce increasingly sophisticated and reliable phonetic studies, there is some hope that we can return to this question a few years down the road.

[12] In essence, F2 is lowered not only by greater backness in tongue position, but also by lip rounding. This introduces a confounding variable to explanations for relative F2 values, such that in a contrast between rounded and unrounded vowels, even if the tongue position is held constant, the F2 of the unrounded vowel will be higher, making it plot acoustically as more central. Thus, acoustic measures do not allow us to identify what portion of the change in F2 is due to rounding and what portion to tongue advancement.

But even if we were to have incontrovertible phonetic data for all languages, the degree of precision of the descriptions of the modern languages (Kari'nja, Garifuna, Tiríyo) is only marginally relevant to the phonetic value of the proto-vowel. Original values may survive, but they may undergo change as well.

In favor of /ë/, there are a number of modern descriptions that place the modern reflex of *ɨ as an unrounded high **back** vowel, including for Kari'nja (Hoff 1968), Hixkaryana (Derbyshire 1985), and Garifuna (where the unrounded back vowel alternates between high and mid). If the unrounded high vowel were to reconstruct as back rather than central, then by analogy this would make more likely the reconstruction of the unrounded mid vowel as back. But rather than rely solely on numbers, in favor of this position we can add an argument that follows from the phonetic properties that would create a disposition for the mergers with four other vowel phonemes, resulting in its ultimate disappearance. Starting from the unrounded mid back vowel /ë/, all four mergers could come about by the change of just one phonetic feature: lowering produced /a/, raising produced /ɨ/ or /ï/, fronting produced /e/, and rounding produced /o/. To realize the merger with /a/ both vowels presumably are equally suited: lowering will do. For the merger with /ɨ, ï/, both vowels are equally well-suited (assuming parallism between the high and mid unrounded vowels): the central /ə/ would raise to the central /ɨ/ or the back /ë/ would raise to the back /ï/. For the merger with /e/, both vowels need to move in the same direction: forward. Our impression is, however, that fronting of back vowels is a more natural change than fronting of central vowels. For the merger with /o/, the back vowel only has to lose its atypical unroundedness. The central vowel, on the other hand, has both to acquire rounding and to move backwards: a complicated change, and therefore less likely to end up in a disappearance by merger (Labov 1994:327-329). This fact puts *ə at a disadvantage, except in those languages where the merger with /o/ did not happen and instead /ə/ is found. For these languages the other candidate, *ë, needs an extra change from back to central. This change, however, is less complicated than the combined change (backing, rounding) needed to move from *ə* towards /o/.

The simplicity of the four changes, each involving only a single feature, should increase the likelihood that the changes would result in mergers, as did happen—rather than in the emergence of new vowels, as otherwise might have happened. For in another respect the conditions for the preservation of the phonological distinction were favorable: there was ample room in phonetic space for distinctions like /e/ versus /ɛ/, /o/ versus /ɔ/, /a/ versus /ɑ/ (cf. Labov 1994:327-329). So we conclude that *ô certainly was an unrounded mid vowel, more likely back than central.

APPENDIX 1. Some correspondences for Proto-Cariban *ô

	about.ERG		arrow2		flea	go	high	IMPER	leave	moon	kill	toad		seize	star	snake		stone	woman	bee/ honey	(*j̈ô tooth)	*jô bone	*j̈ô) cook
	σ_1	σ_2	σ_2	σ_1	σ_2	σ_1	σ_2	σ_1	σ_1	$\sigma_{1/3}$	σ_1	σ_1	σ_2	σ_2	σ_3	σ_1	σ_2	σ_1	σ_1	σ_2	σ_1	σ_1	σ_1	σ_1
Tiriyó	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	e	e	e
Akuriyó	ə	ə	ə	ə	ə	ə		ə	ə		ə				ə			ə	ə			e	e	
Karihona	ə	ə	ə	ɛ̲	ə	ə	ə	ə	ə	ə	ə			ə	ə	ə	ə	ə	ə			e	e	e
Wayana	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	e	e	e
Ye'kwana	ə	ə	ə			ə	ə	ə		ə	ə			ə		ə	ə		o̲			e		
Kari'nja	o	o	i̲	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	e	e	
Island Carib (Breton)	oü̲i	Ø		e	e			éu	oú	um	é			o	Ø			é	é			é	e, a	
Island Carib (Hoff)	ui	Ø		ï	ï			ï	u	ï	ï			o	Ø			ï	ï			e	e, a	
Garifuna				ï											ï			ï	ï					
Apalaí	o	o	o	o	o	o	e	o	o	o	o	o	o	o	ua̲	o	o	o		o	e	e	e	e
Waiwai	o	o		o	o	o	Ø	o	o	i̲	o	o	o	Ø	o	o	o	o		u̲	o	o	o	o
Hixkaryana	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
Katxuyana	o	o		o	o	o	o	o	o		o	o	o	o	o	o	o	o	o	o	o	o	o	o
Tamanaku	o/a	o/e	e	e	e/a	e	e	e		a	e/a	eu	Ø	o	e/a	a	e	e		e	e	e		
Kumana			e	e	e/a	e/o					e/a	eu/o	e/o		e	e/o/a	e/o/i	o	a	e	e	e	e	
Yabarana/ Mapoyo	ə	ə/e̲	é̲	ə	ə	ə/e̲	ə			ə	ə	ə	Ø	ə	ə	ə	ə	ə	ə	ə	ə	e̲/ə	ə	
Panare	ə	Ø		ə	ə	ə	ə	ə		ə	ə			ə	ə	a̲	e̲	o̲	i̲	ə	ə	ə	ə	
Pemón	ə	Ø			ə		ə											o						
Kapong	ə	Ø	ə	ə	ə	ə	Ø	ə	u		ə	ə	ə	Ø/i		ə	ə	ə	Ø	ə	ə			
Makushi	i	Ø	i		i	i	i	i	i		i	i	e	i	i	i	i	i	i		e	e		
Bakairi	ə	ə	ə	ə	ə	i	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə	ə		ə	e	Ø		
Arara	o	Ø	o	a	o				o		o			o		o	o			Ø	e	Ø		
Ikpéng	o	Ø	o	o	o	Ø	o	o	o	o	o	u	o		Ø	o	o			Ø	e	Ø		
Kuhikuru	e	e	e	e	e	e	e	e		e	e			e		e	e	e	Ø	e	i	Ø		
Yukpa	o	o	o		o	e	o	o	o		o			u	u				o	o	o	i	o	
Waimiri Atroari	i	i	u		i	i	i	i		e	u	i	i	e	i			o	e	i	e	i		

Appendix 2. Cognate sets for Proto-Cariban *ô[13]

	about/ERG	arrow-2	flea	go	high	IMPER	leave(TR)	moon	shoot, kill	toad
P.Cariban	*pô̲kô̲	* pɨrô̲w(a)	*tikô̲	*[w]tô̲[mô̲]	*kawô̲	*-kô̲	*nô̲[mô̲]	*nunuwô̲	*[ti]wô̲[nô̲]	*pô̲rô̲ru
P.Taranoan	*pə̲kə̲	*pɨrə̲u	*ʧikə̲	*tə̲[mɨ̲]	*kawə̲	*-kə̲	*nə̲[mɨ/ə]	*nunnə̲	*[ti]wə̲	
Tiriyó	pə̲(kə̲)	pɨrə̲u	sikə̲	tə̲, tə̲mɨ̲	kawə̲, kae̲	-kə̲	nə̲	nunnə̲	[ti]wə̲	pə̲rə̲ru
Akuriyó	pə̲kə̲	pɨrə̲u	ʧikə̲	tə̲, tə̲mɨ̲		-kə̲	nə̲		wə̲	
Karihona	pə̲kə̲	hərə́u	ʧikə̲	tə̲, də̲mɨ̲	kawə̲	-gə̲	nə̲	nunə̲	[ti]wə̲	
Wayana	pə̲k(ə̲)	pɨrə̲u	sikə̲, sihkə̲	tə̲, tə̲m	kawə̲	-k, -kə̲	nə̲mə̲	nunuwə̲	[tu]wə̲	pə̲rə̲
Ye'kwana	hə̲kə̲			tə̲, tə̲mə̲	kawə̲:	-kə̲		nunə̲	tɨwə̲	
Kari'nja	po̲:ko̲	pɨ̈rɨ̈:wa	si:ko̲	wɨ̈to̲	ka:wo̲	-ko̲	no̲:	nu:no̲	wo̲	po̲ro̲:rɨ̈
IC (Breton)	bo̲üic	boulɛ́üa	chicke̲	ni-te̲-m		-ké̲u	no̲úta	nónu̲m	ouè	
IC (Hoff)	/po̲ik/	/purɨ̲ua/	/sikɨ̲/	/tɨ̲/		/-kɨ̲/	/nu̲ta/	/nunɨ̲/	/uɨ̲/	
Garifuna			sigɨ							
Apalaí	po̲ko̲	pɨro̲u	ʃiko̲	to̲	kae̲	-ko̲	no̲mo̲	nuno̲	wo̲	po̲roru
Waiwai	ɸo̲ko̲		si:ko̲	to̲, to̲mo̲	kaw, kaji̲	-ko̲	no̲m	nu:ñi̲	[t]wo̲	ɸo̲ro̲rɨ
Hixkaryana	ho̲ko̲		sɨko̲	to̲	kawo̲	-ko̲	no̲m	nuno̲	[t]wo̲	ho̲ro̲ru
Katxuyana	ho̲ko̲		sɨko̲	to̲, to̲mo̲	kawo̲	-ko̲	no̲mo̲		[t]wo̲	po̲ro̲rɨ
Tamanaku	pa̲ke̲, po̲ko̲-no	pre̲u	ʧike̲	te̲, ta̲	kawe̲	-ke̲		nuna̲	wo̲	peuru
Kumana		pure̲u, pre̲u	chike̲, chiko̲n	ta̲, te̲	kawe̲, kao̲				wa̲, we̲	pue̲re̲r, po̲ro̲ro
Yabarana/ Mapoyo	pə̲ke̲, pə̲kə̲	ré̲u	ʧi:kə̲	tə̲	ka:ə̲, kawe̲	-kə̲		nu:nə̲	wawə̲, wa̲, we̲	pə̲ru
Panare	pə̲ʔ		ʧikə̲	tə̲	kawə̲	-kə̲		wə̲:nə̲	wə̲	
Pemón[2]	pə̲k	(pere̲u)	(chike̲)	tə̲		-kə̲	(ne̲ma̲)		(we̲)	(pe̲re̲teku)
Kapong	pə̲k, bə̲k	pərə̲u	ʧi̲gə̲	də̲		-kə̲, -gə̲	nu̲mɨ̲		wənə̲, wə̲	pə̲rə̲dugu
Makushi	pɨ̲ʔ	pɨrɨ̲u		tɨ̲	kawɨ̲ne	-kɨ̲	nɨ̲mɨ̲		wɨ̲	pɨ̲re̲tiku
Bakairi	wə̲gə̲	pɨrə̲u	ʃigə̲	də̲, tə̲	kaɨ̲	-kə̲, -gə̲	iə̲̃	nunə̲	iə̲	pə̲rə̲ru
Arara	po̲k	pɨro̲m	ʧiga̲	ɨdo̲			ino̲		wo̲	
Ikpéng	wo̲k, po̲k	pɨro̲m	ʧigo̲ŋ	ero̲	ɨkap	-ko̲, -k	ino̲	nuno̲	wo̲	puro̲n

[13] Pemón words in parentheses are taken from Armellada and Salazar (1982), which recognizes neither ɨ nor ə. The other words in the Pemón row are taken from Gildea's field notes with the Arekuna dialect of Pemón.

APPENDIX 2. Cognate sets for Proto-Cariban *ô (continued)

	about/ERG	arrow-2	flea	go	high	IMPER	leave(TR)	moon	shoot, kill	toad
Kuhikuru	hẹkẹ, pẹkẹ	hiɣẹ	sikẹ	tẹ	kapẹhe	-kẹ		ŋunẹ	ẹ, hẹ	
Yukpa	po, pokọ			to	kajẹ	-ko, -k	ejo		wọ	
Waimiri Atroari	pɨkɨ	pɨrụwa		tɨn	kawɨ	-kɨ		nenuwẹ	wụ	pɨrɨrɨ

	seize	star	snake	stone	woman	bee/honey	tooth	bone	cook
P.Cariban	*apôti	*tirikô	*ôkôju	*tôpu	*wôriti	*wanô	*jô	*jôtipô/ɨ	*[t-]jô
P.Taranoan	*apətʃi	*tʃirikə	*əkəi	*təpu	*wər(i/i)tʃi		*jẹ	*jẹtipə	
Tiriyó	apəi	sirikə	əkəi	təpu	wəri	wanə	je	jetipə	[t]je
Akuriyó		sirikə		təpu	wəriʔi		jeʔ	jeʔpə	
Karihona	apətʃi	tʃirikə	əkəi	təhu	wəritʃi		je	jetihə	je
Wayana	apəi	sirikə	əkəi	təpu	wərii	wanə	je	jetpɨ	[t]je
Ye'kwana	ahəi		əkə:jɨ		wọri		ðẹ		
Kari'nja	apo:i	siri:ko	oko:ju	to:pu	wo:rĩi	wa:no	je	je:po	
IC (Breton)	abọüi	chiric_		tébou	oüélle		iéri	epou, abo	
IC (Hoff)	/aboi/	/sirik/		/tïbu/	/uïri/		/ieri/	/epu, abu/	
Garifuna		sirigï		dïbu	uïri				
Apalaí	apoi	ʃirikua(:)to	okoi	topu		ano	je, zẹ	jeʔpɨ, zẹʔpɨ	ẹ
Waiwai	ahsɨ	ʃirʲko	okoji	to:ɸu		we:nu	jo	jotʃhɨ	[t]jo
Hixkaryana	ahosɨ	ʃerʲko	okoje	tohu	woriskomo	weno	jo	jotʃhɨ	[ti-]jo
Katxuyana	ahosɨ	tʃir(ʲ)ko	okoji	tohu	worisɨ	weno	jo	jotʃpɨ	jo
Tamanaku	apoiʔ	tʃirika, tʃirikẹ	akẹi	tẹpu		wanẹ	je	jetpẹ	
Kumana		tʃirkẹ	ẹkẹi, aki, okoju	topo	wariz, waritʃe	wánẹ, wanẹ	ẹ	jepo, jep, tʃẹp	
Yabarana/ Mapoyo	apəti?	tʃirikə, sihkə	əkəi	tə:pu	wəria'mu	wa:nə, wánə	éma jə, je	jəhpə, jəʔbə	
Panare	apəsɨ	tjakəŋ	akẹ, aʔkẹ	toʔ	wɨnkiʔ, wɨnkit	wanə, wənə	jə	əhpə	
Pemón	(apichi)	(chirikẹ)	(okoi)	toʔ	(weri)	(wan)	(yẹ)	(yẹpue)	
Kapong	aʔtʃi, apisi		əgəi	tək, təbu	uriʔtʃaŋ		jə, ə	əʔpɨ	

Appendix 2. Cognate sets for Proto-Cariban *ô (continued)

	seize	star	snake	stone	woman	bee/honey	tooth	bone	cook
Makushi	apɨʔsi	sirikɨ	ɨkɨi	tɨʔ	wɨriʔ		je, e	jɛʔpɨ	
Bakairi	awə	ʃimukə	əgəu	təu		pɛ̃rɛ̃	ɛ	ibirɨ	
Arara	aboɨ		oko̱i			wan	ɛ	itpɨ-n	
Ikpéŋg		tiriŋ	ogoj			paŋ	ɛ	itpin	
Kuhikuru	ihe		eke	tehu	itâõ	iŋe	i	ipɨɣi	
Yukpa	apuje	ʃeku		top	worepa	wano	ɨ	jowi	
Waimiri Atroari	apeʔ	ʃirkɨ, ʃiriki		topi	weri	wanɨ	jeː	jihi	

Appendix 3. Cognate sets that show (unsystematic) word-initial *ô > a[14]

	path	who	what	I	you	food (unpossessed)	food (possessed)
P.Cariban	*ôtema	*ônôkɨ	*ôtɨ	*ôwɨ-rô	*ômô-rô	*ôrepa	*ɨ-ôrepa-rɨ
P.Taranoan	*ətʃema, etʃema	*ənəkɨ	*ətɨ	*əwɨ	*əmə(rə)	*ərepa	*erepa
Tiriyó	əema, e:ma	**a**kɨ	**a**tɨ	wɨ[:]	əmə	ərepa	erepa
Akuriyó	əʔema	əkɨ	ətɨ	wɨ	əmə		erepa
Karihona	esema	ənəkɨ	ətɨ	əwɨ	əmərə	əreha	ereha
Wayana	əhema, ehema	ənɨk	ətɨ	ɨwu	əmə		erepa
Ye'kwana				ɨwɨ	əmədə		
Kari'nja	e:ma	**a**no:kï, no:kï	o:tï	**a**u, **a**:u	**a**mo:ro	arepa	erepa-rɨ
IC (Breton)	éma	**a**naki	**a**tt-[3]	**a**o	**a**mánle	aléiba	
IC (Hoff)	/ïma/	/anaki/	/at-/	/au/	/amïrï/	/arepa/	
Garifuna	ïma			**a**u	**a**mïrï		
Apalaí	osema	onoki	otɨ	ɨwɨ	omoro		
Waiwai	esama	onoke	**a**ti, **a**htʃe	owɨ	omoro		erewsɨ
Hixkaryana	**a**sama, osama	onoki	eteni	uro	omoro		erewsɨ
Katxuyana	esma	na:kɨ	ti:mo	owɨ	omoro		
Tamanaku	**a**tʃema	**a**nek, **a**neʔ		ure	**a**mare	arepa	
Kumana	**a**zema, etʃema	**a**nek, enеke	eti, et	ure, utxe	**a**muere		erepa
Yabarana/ Mapoyo	e:ma, **a**tʃema	nəkki, **a**neki	əti, **a**ti, **a**tti	urə, wərə, wihɨ	mo(re), məːrə		
Panare	ʧama, ʧima	nəʔ	əh-kaiʔ, əh-tə	ju	**a**məŋ		
Pemón	(ema)	(ane)	(ek)	(j)urə	**a**mare/**a**ma:rɨ		
Kapong	eʔma	ənik, ənɨʔ	əʔ-rə	(j)urə	**a**mə(rə)		
Makushi	eʔma	**a**ni, **a**niʔ	ɨʔ	u:rɨ	**a**mɨrɨ		
Bakairi	ɜ̃wa, õwã	ɜ̃gɨ	ədɨ	urə	əmə		
Arara	**a**nma	nok	odɨ, **a**te	uro			
Ikpéng	**a**nma	onok	**a**rɨ	uro	omro, omo		
Kuhikuru	**a**ma, ima		tɨ	uye	e(:)ye		
Yukpa	osema	no		**a**wɨ	**a**mo, **a**mor		
Waimiri Atroari	wuma, k-ɨma			**a**wɨ	**a**mɨ, **a**mɨrɨ		

[14] 'What' in IC (Breton) is limited to the fixed expression *attoüati* 'yes, there is nothing against it'.

Appendix 4. Language sources for cognate sets

Language	Source
P.Cariban	Gildea & Payne 2007, Meira 2002
P.Taranoan	Meira 2000
Tiriyó	Meira 2000
Akuriyó	Meira 2000
Karihona	Meira 2000
Wayana	Tavares 2005, Meira's field notes, Tavares' field notes
Ye'kwana	Hall 1988, Mattéi-Muller & Henley 1990, Cáceres p.c. with Gildea
Kari'nja	Hoff 1968, Mosonyi 1978, Hoff's field notes
IC (Breton)	Breton 1655 (1999)
IC (Hoff)	Taylor 1977, Taylor & Hoff 1980, Taylor p.c. with Hoff
Garifuna	Taylor 1951, Taylor p.c. with Hoff
Apalaí	Koehn & Koehn 1986, 1995; Gildea field notes, Meira field notes
Waiwai	Hawkins 1998, Gildea field notes, Meira field notes
Hixkaryana	Derbyshire 1961, 1979, 1985; Gildea field notes, Meira field notes
Katxuyana	Gildea field notes, Meira field notes
Tamanaku	Gilij 1680-84 (1965), Mattéi-Muller & Henley 1990, Meira & Gildea in prep.
Kumana	Mattéi-Muller & Henley 1990
Yabarana/ Mapoyo	Mattéi-Muller 2004, Mattéi-Muller pc with Gildea
Panare	Mattéi-Muller 1994, Payne & Payne 1999, Gildea field notes
Pemón	Armellada & Salazar 1982
Kapong	Fox 2003, Gildea's field notes
Makushi	Abbott 1990, Amodio & Pira 1996
Bakairi	Meira & Franchetto 2005, Meira's field notes
Arara	Meira & Franchetto 2005, Souza, I. 1988, Souza S. 1993
Ikpéng	Campetela 1997, 2002; Pacheco 1997, 2001, Pacheco p.c. with Gildea, Gildea's field notes
Kuhikuru	Meira & Franchetto 2005, Santos 2005
Yukpa	Meira & Franchetto 2005, Meira's field notes
Waimiri Atroari	Bruno 2003, Lacerda & Queiroz 1991, Bruno p.c. with Meira

References

Abbott, Miriam. 1991. Macushi. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, vol. 3, 23-160. Berlin: Mouton de Gruyter.

Adam, Lucien. 1893. *Matériaux pour servir à l'établissement d'une grammaire comparé des dialectes de la famille caribe* (Bibliothèque Linguistique Américaine tome 17). Paris: Maissonneuve.

Ahlbrinck, Wilhelmus G. 1931. *Encylopedie der Karaïben* (Behelzend taal, zeden en de gewoonten de zer Indianen, Bd. 37, No. 1). Amsterdam: Verh. Ak. Wetenschappen, Afd. let., NF.

Álvarez, José. 2000. La entrada verbal en la lexicografía del Pemón (Caribe). *Boletín de Lingüística* 15. 65-94.

Amodio, Emanuele & Vicente Pira. 1996. *Língua Maxixi-macusi maimu: Guia para o aprenzizado e discionário de lingual macuxi.* Boa Vista (Roraima): Diocese de Roraima and Missionários de Scarboro.

Armellada, Cesário de & Mariano Gutiérrez Salazar. 1981. *Diccionario Pemón*. Caracas: Ediciones Corpoven.

Biet, Antoine. 1664. *Voyage de la France équinoxiale en l'Isle de Cayenne*. Paris: François Clouzier.

Boomert, Arie. 1995. Island Carib archeology. In Neil L. Whitehead (ed.), *Wolves from the sea*, 23-36. Leiden: KITLV Press.

Boyer, Paul. 1654. *Véritable relation de tout ce qui s'est sait et passé au voyage que Monsieur de Bretigny fit à l'Amérique occidentale*. Paris: Pierre Rocolet.

Breton, Raymond. 1655. *Dictionnaire Caraïbe-Français*. Auxerre: Gilles Bouquet. Nouvelle édition 1999. Paris: Karthala et IRD.

Bruno, Ana Carla dos Santos. 2003. *Waimiri Atroari grammar: Some phonological, morphological and syntactic aspects*. Tucson, AZ: University of Arizona dissertation.

Caesar-Fox, Desrey. 2003. *Zauro'nödok Agawaio yau: Variants of Akawaio spoken at Waramadong*. Houston, TX: Rice University dissertation.

Campetela, Cilene. 1997. *Análise do system de marcação de caso nas orações independents da língua Ikpeng.* Campinas, Brazil: Universidade Estadual de Campinas MA thesis.

Campetela, Cilene. 2002. *Aspectos prosódicos da língua Ikpeng*. Campinas, Brazil: Universidade Estadual de Campinas dissertation.

Carlin, Eithne B. 2004. *A grammar of Trio, a Cariban language of Suriname* (Duisburg Papers on Research in Language and Culture 55). Frankfurt-am-Main: Peter Lang.

de Armellada, Cesário & Mariano Gutiérrez Salazar (see Armellada)

de Goeje, Claudius H. (see Goeje)

de Souza, Isaac Costa (see Souza)

de Souza, Shirley Dias Cardoso (see Souza)

Derbyshire, Desmond C. 1961. Hixkaryana (Carib) syntax structure. *International Journal of American Linguistics* 27. 125-142 & 226-236.

Derbyshire, Desmond C. 1965. *Textos Hixkaryâna* (Publicações Avulsas 3). Belém: Museu Goeldi.

Derbyshire, Desmond C. 1979. *Hixkaryana* (Lingua Descriptive Studies 1). Amsterdam: North Holland.

Derbyshire, Desmond C. 1985. *Hixkaryana and linguistic typology*. Arlington: Summer Institute of Linguistics & the University of Texas at Arlington.
dos Santos, G. Mara Ferreira (see Santos)
Edwards, Walter F. 1978. A preliminary sketch of Arekuna (Carib) phonology. *International Journal of American Linguistics* 44. 223-227
Gildea, Spike. 2003. The Venezuelan branch of the Cariban language family. *Amérindia* 28. 7-32.
Gildea, Spike & Doris L. Payne. 2007. Is Greenberg's Macro-Carib viable? In Ana Vilacy Galucio & Pieter Muysken (eds.), *Lingüística histórica na América do Sul* (Boletim do Museu Emilio Goeldi, Série de Ciências Humanas). Belém: Museu Goeldi.
Gilij, Filippo Salvatore. 1780-1784. *Saggio di storia americana*, vols. 1-3. Rome: Publisher.
Girard, Victor. 1971. *Proto-Carib phonology*. Berkeley, CA: University of California dissertation.
Goeje, Claudius H. de. 1909. *Études linguistiques Caraïbes*. Amsterdam: Verhandelingen der Koninklijke Academie van Wetenschappen te Amsterdam.
Goeje, Claudius H. de. 1946. *Études linguistiques Caribes,* tome II. Amsterdam: Verhandelingen der Koninklijke Nederlandsche Academie van Wetenschappen.
Hall, Katherine. 1988. *The morphosyntax of discourse in De'kwana Carib*. St. Louis, MO: Washington University dissertation.
Hawkins, Robert. 1998. Waiwai. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, vol. 4, 3-202. Berlin: Mouton de Gruyter.
Hoff, Berend J. 1968. *The Carib language: Phonology, morphonology, morphology, texts and word index* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land-, en Volkenkunde 55). The Hague: Nijhoff.
Hoff, Berend J. 1995. Language contact, war, and Amerindian historical tradition: The special case of the Island Carib. In Neil L. Whitehead (ed.), *Wolves from the sea*, 37-60. Leiden: KITLV Press.
Hoff, Berend J. 2000. Klemtoon en accent in het Caribisch van Suriname. Unpublished ms.
Jones, Daniel. 1960. *An outline of English phonetics,* 9th edn. Cambridge: Heffer.
Koehn, Edward & Sally Sharp Koehn. 1986. Apalai. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, vol. 1, 33-127. Berlin: Mouton de Gruyter.
Koehn, Edward & Sally Sharp Koehn. 1995. *Vocabulário básico Apalaí-Português: Dicionário da língua Apalaí*. Cuiabá: Summer Institute of Linguistics.
Labov, William. 1994. *Principles of linguistic change, volume 1: Internal factors*. Oxford & Cambridge, MA: Blackwell.
Lacerda, Edith & Carlos Augusto Queiroz. 1991. *Waimiri Atroari preliminary dictionary.* Manaus: PWA.
Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago & London: University of Chicago Press.
Ladefoged, Peter. 2005 (2001). *Vowels and consonants,* 2nd edn. Oxford: Blackwell.
Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Mattéi-Muller, Marie-Claude. 1994. *Diccionário ilustrado Panare-Español, índice Español-Panare.* Caracas: Comisión Quinto-Centenário.

Mattéi-Muller, Marie-Claude. 2002. En busca de criterios clasificatorios para la familia caribe. Linguas Indigenas Brasileiras: Fonologia, gramática e história. In Ana Suelly Arruda Câmara Cabral & Aryon Dall'igna Rodrigues (eds.), *The 1st Encontro Internacional do Grupo de Trabalho sobre línguas Indígenas da ANPOLL,* 37-53. Belém: Editora Universitária, Universidade Federal do Pará.

Mattéi-Muller, Marie-Claude. 2003. Pémono: Eslabón perdido entre Mapoyo y Yawarana. *Amérindia* 28. 33-54.

Mattéi-Muller, Marie-Claude & Paul Henley. 1990. *Los Tamanaku: Su lengua, su vida*. San Cristobal: Universidad Católica del Táchira.

Meira, Sérgio. 1999. *A grammar of Tiriyó*. Houston, TX: Rice University dissertation.

Meira, Sérgio. 2000. *A reconstruction of Proto-Taranoan: Phonology and morphology.* Munich: LINCOM Europa.

Meira, Sérgio. 2002. A first comparison of pronominal and demonstrative systems in the Cariban language family. In Milly Crevels, Simon van der Kerke, Hein van der Voort & Sérgio Meira (eds.), *Current studies in South American languages* (Indigenous Languages of Latin America Series 3), 255-275. Leiden: Research School of Asian, African, and Amerindian Studies.

Meira, Sérgio. 2003. A marcação de pessoa nos verbos em Bakairi (Karíb). In Francesc Queixalós (ed.) *Ergatividade na Amazônia II*. Paris: CNRS-CELIA.

Meira, Sérgio & Bruna Franchetto. 2005. The southern Cariban languages and the Cariban family. *International Journal of American Linguistics* 71. 127-192.

Meira, Sérgio, Spike Gildea & Berend J. Hoff. Forthcoming. On the origin of ablaut in the Cariban family. In Spike Gildea & Ana Vilacy Galucio (eds.), *Historical linguistics in South America*, special issue of the *International Journal of American Linguistics*.

Mosonyi, Jorge C. 1978. *Diccionario basico del idioma Cariña*. Caracas, Venezuela: Universidad Central de Venezuela dissertation.

Pacheco, Frantomé. 1997. *Aspectos da gramática Ikpeng.* Campinas, Brazil: Universidade Estadual de Campinas MA thesis.

Pacheco, Frantomé. 2001. *Morfossintaxe do verbo Ikpeng.* Campinas, Brazil: Universidade Estadual de Campinas dissertation.

Payne, Thomas & Doris L. Payne. 1999. Panare: A Cariban language of central Venezuela. Unpublished ms.

Pelleprat, Pierre. 1655. *Relation des missions des PP. de la compagnie de Jésus dans les isles, et dans la terre ferme de l'Amérique Méridionale, avec une introduction à la langue des Galibis*. Paris: Cramoisy.

Santos, G. Mara Ferreira dos. 2002. *Morfologia Kuikuro: As categories 'nome' e 'verbo' e os processos de transitivização e intransitivização.* Rio de Janeiro, Brazil: Universidade Federal de Rio de Janeiro MA thesis.

Santos, G. Mara Ferreira dos. 2005. The flexional morphological classes in the Kuikuro language. Paper presented at Grammaire des Langues caribes - Grammar of Cariban Languages. Paris.

Souza, Isaac Costa de. 1988. *Contribuição para a fonologia da língua Arara (Karib)*. Campinas, Brazil: Universidade Estadual de Campinas MA thesis.

Souza, Shirley Dias Cardoso de. 1993. *Alguns aspectos morfológicos da língua Arara (Karib)*. Brasília, Brazil: Universidade de Brasília MA thesis.
Souza Cruz, Odileiz. 2005. *Fonología e gramática Ingarikó*. Amsterdam, The Netherlands: Free University of Amsterdam dissertation.
Tavares, Petronila. 2005. *A grammar of Wayana*. Houston, TX: Rice University dissertation.
Taylor, Douglas. 1951. *The Black Carib of British Honduras* (Viking Fund Publications in Anthropology 17). New York: Wenner Gren Foundation. 1963, Johnson Reprint.
Taylor, Douglas. 1977. *Languages of the West Indies*. Baltimore & London: The Johns Hopkins University Press.
Taylor, Douglas & Berend J. Hoff. 1980. The linguistic repertory of the Island-Carib in the 17th century: The men's language — a Carib pidgin? *International Journal of American Linguistics* 46. 301-312.

Spike Gildea
spike@uoregon.edu

B.J. Hoff
B.J.Hoff@hum.leidenuniv.nl

Sérgio Meira
s.meira@let.leidenuniv.nl

6

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 125-145
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4453

Multiple functions, multiple techniques: The role of methodology in a study of Zapotec determiners[1]

Donna Fenton

University of California, Berkeley

Field linguists use a combination of techniques to compile a grammatical description, starting with various types of targeted elicitation and followed by the study of more natural speech in the form of recorded texts. These usual techniques were employed in my work on Teotitlán del Valle Zapotec, an Oto-Manguean language spoken in Mexico, but in an unusual order, with texts, mainly folk tales and life histories told by community elders, being collected and analyzed first due to the priorities of the documentation project I was a part of. This paper examines the role that methodology played in the investigation into one small area of the grammar, a set of noun phrase-final determiner clitics. These clitics make both spatial and temporal distinctions, raising theoretical questions regarding the role of a temporal marker in the NP. At the same time, it brought to light some interesting issues surrounding methodology in fieldwork: how does the method of collection affect the type of data gathered, and does the order in which different methodologies are employed affect the final outcome?

1. INTRODUCTION. While transcribing and translating Teotitlán del Valle Zapotec texts as part of a documentation project, one of the initial puzzles I encountered was a set of noun phrase-final clitics for which the interpretations of native speaker consultants varied widely. Further investigation, consisting of targeted elicitation and comparison with related dialects, revealed the clitics to be determiners which make both spatial and temporal distinctions, and which have important discourse functions. This investigation raised theoretical questions regarding the role of a temporal marker in the NP. At the same time, it brought to light some interesting issues surrounding methodology in fieldwork: how does the method of collection affect the type of data gathered, and does the order in which different methodologies are employed affect the final outcome?

2. LANGUAGE BACKGROUND AND METHODOLOGY. Teotitlán del Valle Zapotec (TdVZ) is a largely undocumented variety of Central Zapotec (Zapotecan, Otomanguean) spoken in Teotitlán del Valle, a town located in the Tlacolula branch of the Central Valley

[1] I would like to express my gratitude to Troi Carleton for giving me the opportunity to do fieldwork in Teotitlán del Valle, and to Troi and several other project participants who were generous in sharing their data and insights during our many discussions in the field that contributed to this research. Many thanks also to Lynn Nichols for her guidance and helpful feedback on earlier drafts of this paper.

System of Oaxaca State, Mexico. The 2005 census reports a population of about 5600 people, about two-thirds of whom are reported to be "speakers of some indigenous language" (INEGI 2005).[2] Smith Stark (2003) classifies TdVZ as one of the dialects that make up the group Western Tlacolula Zapotec, identified by the code 'zab' in the Ethnologue (Gordon 2005). Recent work on other Western Tlacolula Zapotec dialects includes Munro & Lopez (1999) and Lillehaugen (2006). TdVZ is an endangered language, but some children are learning it as a first language and there is community interest in language preservation and revitalization.

My work on TdVZ began in 2004 with my participation in the Teotitlán del Valle Community Language Archive Project, a collaboration between San Francisco State University and the Community Museum of Teotitlán del Valle under the direction of Professor Troi Carleton. Because the main interest of the museum committee was the creation of an archive of oral histories and local folklore told by elders, work on text collection, transcription and translation began almost immediately, after a basic phonemic inventory had been established but before substantial work had been done on the grammar of the language. A small amount of elicitation of word lists and paradigms was done along with the collection of texts during the initial stages, but the bulk of my elicitation and grammatical analysis was carried out during my fieldwork in Teotitlán del Valle in the summer of 2007.[3]

Field linguists generally use a combination of targeted elicitation and texts to compile a grammatical description, with texts providing more naturally occurring forms free from the biases of the elicitation situation, and with elicitation being useful for providing full paradigms along with native speaker grammaticality judgments (Mithun 2001, Chelliah 2001). In my investigation of the semantics of the TdVZ determiner clitics, both elicitation and the analysis of texts have played vital roles. The texts, which owing to the circumstances of my work in Teotitlán del Valle were collected and transcribed in the early stages of the research, revealed the temporal and discourse functions of the determiners. Elicitation, which was informed by the analysis of the texts, brought out the expected spatial distinctions and enabled me to refine my analysis of the temporal extensions of the determiners. Neither method on its own would have provided a complete picture, and it is possible that the 'backwards' order of data collection allowed for a better insight into how speakers actually used and thought about the system of determiners. Had I known going into the text transcription that *=ki* is 'simply' the distal determiner, I may have glossed it as such and moved on without having the interesting discussions with consultants that prompted the investigation. Thus native speaker intuitions, obtained during elicitation and translation of the texts, also play an important role in my analysis.

In most cases Spanish was the contact language used for elicitation and text translation, and the glosses were subsequently translated into English; I have included both the original Spanish and the English free translations in my examples. For examples from the

---

[2] Almost all of them would be speakers of TdVZ; there are no significant numbers of speakers of any other indigenous language in the town. I have heard of a few speakers of other Zapotec languages who moved to the town as adults.

[3] 2007 fieldwork funded by UC Berkeley Graduate Division Summer Research Grant.

few elicitation sessions conducted in English, I have only included the English free translation. There is a great deal of speaker variation in the use of tone and phonation types in TdVZ, and these have yet to be fully analyzed. I have marked the low/creaky vowel on the proximal determiner with the grave accent, as used in the practical orthography. Otherwise, tone is not marked in the examples.[4] Glottalized vowels are marked with an apostrophe. The consultants who made this research possible are Lorenzo Gutiérrez Lopez (LGL), Serafin Martinez Gonzales (SMG), Samuel Bautista Lazo (SML), Lucia Lazaro Vasquez (LLV), Viridiana Chavez (VC), Andrés Lazaro Bautista (ALB), and Maria Mendoza Contreras (MMC).

3. TDVZ DETERMINER CLITICS: DISTRIBUTION AND FUNCTION. Like several other Zapotec languages, Teotitlan del Valle Zapotec (TdVZ) has a set of determiners that make rather fine distinctions regarding the spatial relationship between the speaker and the referent. TdVZ has four such determiners, which appear as clitics at the end of the noun phrase; these are shown in Table 1 along with their approximate English and Spanish equivalents.

TABLE 1. TdVZ determiner clitics

	TdV Zapotec	Spanish	English
proximal (PROX)	*rè* /rɐ̰^{L}/	'este'	'this'
medial (MED)	*kang* /kaŋ/	'ese'	'that'
distal (DIST)	*re* /re^{H}/	'ese/aquel'	'that'
distal/invisible (INVIS)	*ki* /ki/	'aquel (ausente)'	'yon (invisible)'

This set of NP-final clitics acts as definite determiners and also indicates the spatial relation between the object denoted by the NP and the speaker; their function overlaps the categories of definite article and demonstrative adjective in English and Spanish. In addition to their deictic spatial reference, these clitics have developed a temporal sense, with the non-visible determiner *=ki* functioning as a past tense marker and contrasting with the present tense sense of the other three. Although it is relatively common for spatial demonstratives to expand into the temporal domain (Givón 2001), in a language like TdVZ where the aspectual prefix on the verb does not necessarily give any indication of the time of the occurrence of the event relative to the time of the utterance, the addition of a tense-like determiner means that the NP carries additional temporal information not encoded in the verb.

Consultants varied widely in their glosses for the enclitics when providing translations for the text, some translating them as corresponding to the Spanish demonstrative adjective or definite article, others calling them discourse markers that have no Spanish translations,

[4] De Angulo (1926) and Lowes & Lopez Cruz (2007) discuss the use of tone in TdVZ verbal morphology.

and most attributing some sort of 'past tense' meaning to the distal *=ki. =ki* was by far the most commonly used of the determiners in the texts I collected, which were folktales, local legends, and personal narratives set in the distant past.

As the following examples show, the determiners appear in NP-final position, attaching directly to the noun they modify (1), following other nominal modifiers (2), or at the end of a relative clause (3). For phonological reasons, I analyze them as clitics; they are always unstressed and seem to form a single unit with the word to which they attach.

(1) (SMG, elicited)[5]
gu-lɛɛbi̯ *nis=rè*
COMPL-boil water=PROX
'his water was boiled'
[$_{NP}$[$_{N}$*nis*] *=rè*]

(2) (SMG, elicited)
gu-daaw-a' *naraʒ* *laa=rè*
COMPL-eat-1S orange bitter=PROX
'comí mi naranja amarga'
'I ate my bitter orange'

(3) (LGL, text: *Pobreza*)
per *bɛtsa-n* *ni* *zuwa'a* *sru'u=ki* *ri-dʒietʃ-an*
but brother-3S REL HAB.sit well=INVIS HAB-be.angry-3S
'pero su hermano, el que vive bien, se enoja'
'but his brother, the one who lived well, got angry'
[$_{NP}$ [$_{RelP}$ *ni zuwa'a sru'u*] *=ki*]

The determiners are used with definite referents, and are in complimentary distribution with the indefinite article *te*, which occurs in prenominal position. *Te* is usually used when a referent is first introduced into the discourse, and the determiner enclitics are used in subsequent references, as illustrated in (4).

---

[5] Abbreviations used: 1S='first person singular'; 1P='first person plural'; 2S='second person singular'; 3S='third person singular'; 3S.ANI='third singular, animal'; 3S.INAN='third singular, inanimate'; 3S.OBJ='third singular object'; 3P='third person plural'; COMPL='completive aspect'; COP='copula'; DIST='distal determiner'; FUT='future aspect'; HAB='habitual aspect'; INT='intensifier'; INVIS='distal/non-visible determiner'; MED='medial determiner'; NEG='negative/negation'; NMLZ='nominalizer'; POT='potential aspect'; PL='plural'; POSS='possessive'; PROG='progressive aspect'; PROX='proximal determiner'; PST='past'; REL='relative pronoun'; TAM='Tense-Aspect-Mood'; TdVZ=Teotitlán del Valle Zapotec; TT=topic time; TU=time of utterance; TSit=situation time.

(4) (LGL, text: "El Tecolote y la Culebra")

tʃi	*bi-ɛnela'aw-an*	*zit^{h}*	*ra-shia-teè*	***te***	***maiŋ***	*aʂ*	*ri-niɛn*
when	COMPL-see-3S	far	HAB-jump-INT	a	animal	and	HAB-say-3S

'Cuando vio lejos un animal brincaba y decía'
'When he saw in the distance an animal jumping, and said,'

tutu	*maiŋ=re*	*ra-shia*	*lo*	*nɛ:z*
what	animal=DIST	HAB-jump	face	road

'"¿que animal será el que brinca en el camino?"'
'"what animal could that be jumping in the road?"'

aʂ	*ɣu-zlo*	*ka-ʃibi-an*	*aʂ*	*neŋ*	*trakwe'edio*	*tu*
and	COMPL-begin	PROG-be.afraid-3S	and	say.3S	who.knows	what

'y empezó a tener miedo y dijo "¿quien sabe que'
'and he began to be afraid, and said, "Who knows what'

maiŋ=kaŋ	*neŋ*	*aʂ*	*ɣu-zlo*	*ka-za'-aŋ*
animal=MED	say.3S	and	COMPL-begin	PROG-walk-3S

'animal es?" dijo y empezó a caminar.'
'animal that is", and he began to walk.'

tʃi	*bi-tsuin^{j}-aŋ*	*ɣut*	*ra-shia*	***maiŋ=ki***	*tʃi*
when	COMPL-arrive-3S	where	HAB-jump	animal=INVIS	when

'Cuando llegó donde brincaba aquel animal,'
'When he arrived where that animal was jumping,'

ɣu-n-aŋ	***te***	***damen***	*kon*	***te***	***beļ***	***bel=ki***
COMPL-see-3S	a	owl	with	a	snake	snake=INVIS

'cuando vio era un tecolote y una culebra. Ese culebra'
'he saw that it was an owl and a snake. The snake'

a-bi-del-teè-m	*le'en*	*ʃiļ̨*	***dam=ki***
already-COMPL-tangled.up-int-3S.ANI	inside	wings	owl=INVIS

a-ba'-an^{j}-um	*ɣan*
already-COMPL-do-3S.ANI	win

'ya enredó dentro de las alas del tecolote, ya ganó.'
'had tangled itself up inside the wings of the owl, he had already won.'

The enclitics are rarely used directly with inherently definite NPs, such as proper names and pronouns, or with generic or mass nouns (as in (5)). They generally do not occur following possessed nouns, unless the speaker needs to disambiguate. However, they

do often occur with relative clauses that contain inherently definite NPs, as in (6); in this case I analyze the determiner as modifying the relative pronoun *ni*.

(5) (LGL, text: *La Bruja Sacó a Erasto*)

jonum	*bi-dɛ'un*	***nopi***	*per*	*bɛ'-ɛt*	***tʃikᵾ***
three	COMPL-drink-1P	**mezcal**	but	COMPL-come	**Francisco**

'Los tres tomamos mezcal y vino Francisco,'
'The three of us were drinking mezcal and Francisco,'

duʃ-kuɲad-uŋ	*kon*	*dunoŋ*
POSS-bro.in.law-1P	with	1P

'nuestro cuñado.'
'our brother-in-law, came.'

(6) (LLV, text: *Historia de Pancho Villa*)

per	*ni*	*la*	*da zikᵾ=ki*	*ɣu-ti-aŋ*
but	REL	named	Don.Ziku=INVIS	COMPL-die-3S

'Pero el que se llama Don Ziku se murió.'
'But the one called Don Ziku died.'

They can also occur after the head of a relative clause, and in some cases, in both positions:

(7) (SMG, elicited)

a.

gu-rap-u'	*ð-tap*	*libr=ki*	*ni*	*gu-ni-ɛ*
COMPL-have-2S	PL-four	book=INVIS	REL	COMPL-say-1S

lui	(*nai*)
you	(yesterday)

'¿Tienes los cuatro libros que te pedí (ayer)?'
'Do you have the four books I asked you for yesterday?'

b.

gu-rap-u'	*ð-tap*	*libr=ki*	*ni*	*gu-ni-ɛ*	*lui=ki*
COMPL-have-2S	PL-four	book=INVIS	REL	COMPL-say-1S	you=INVIS

'¿Tienes los cuatro libros que te pedí?'
'Do you have the four books I asked you for?'

The use of *=ki* in both positions in (7b) adds an emphatic sense. According to SMG, this wording implies that the speaker has already asked for the books many times and so is emphasizing both the books and the asking.

The proximal and distal determiners, *=rè and =re*, are distinguished by tone and phonation type, and are homophonous with the independent words meaning 'here' and 'there', respectively. *=re/re* are marked by high tone, while *=rè/rè* have low tone and creaky voice

phonation.[6] The two sets are obviously closely related semantically; they are distinguished by the environment in which they occur, but there are some cases of ambiguity. The two types can co-occur in the same sentence:

(8) (SMG, elicited)

tu	*ð-bɛni*	*ngiu*	*ni*	*na*	*tjop=re*	*zu*	*re*
who	PL-person	man	REL	COP	two=DIST	HAB.stand	there

'¿Quienes son los dos hombres allá?'

'Who are those two men over there?'

In addition, there are independent demonstrative pronouns, *ndè* 'this (*este*)' and *dè* 'that (*ese*)'. There are also at least two related independent words with anaphoric discourse reference: *reki* 'there, that place (formerly specified)' and *zeki* 'that way/manner (previously elaborated in the discourse)'.

Other Central Zapotec languages also have complex systems of determiners that make fine distinctions in spatial reference. Isthmus Zapotec has the four-way system /ri'/ 'this (*este*)', /ka/ 'that (*ese*)', /rika'/ 'that (*aquel*)', /ke/ 'yon, invisible (*aquel ausente*)' (Pickett & Black 1998). The San Lucas Quiaviní Zapotec Dictionary (Munro & Lopez 1999) lists the independent demonstrative adjectives *re'nn, rèe', rèe, rèenn,* and the demonstrative clitics *=ag* 'PROX' and *=ih* 'DIST'. The demonstrative adjectives in both of these languages appear in the same position as those in TdVZ, and the distal /ke/ in Isthmus Zapotec is used as a temporal marker in the same fashion as its TdVZ cognate *=ki.*

4. ESTABLISHING A SPATIAL SENSE: ELICITATION. In my initial analysis of the texts, there was no indication that these determiner clitics had any spatial reference. Their function in the texts seemed to correlate with the majority of the native speaker intuitions, that they were either temporal or discourse markers: *=ki* always occurred with the description of past events, and the fact that a determiner sometimes failed to appear in the expected environment indicated a possible discourse-level function. Only one consultant gave any indication of a spatial meaning: LGL said that "*=ki* is further away [than *=re*] in time and space," and he also used the Spanish determiners *ese* and *aquel* in his glosses. Because it is common cross-linguistically and within the Zapotec family for determiners to make a spatial distinction, I expected to find that these determiners would also have a spatial sense, and indeed when eliciting phrases with prompts that made spatial distinctions, consultants used the determiner clitics in a very consistent manner. I assume the spatial reference to be the primary sense, as the TdVZ demonstrative determiners have cognates with spatial reference in other Zapotec languages, and also because this type of semantic shift is common, as noted above.

---

[6] With some speakers, the vowel in the proximal expressions is longer; they have the pairs *-reè/reè* versus *-re/re*. One consultant, LGL, also distinguishes the two sets by vowel quality. He pronounces the determiners /ra^{H}/ and /ra^{L}/, and uses the standard vowel quality for the locative adverbial expressions equivalent to English 'here' and 'there'. To avoid confusion, I use the transcription for the more 'standard' pronunciations of the determiners throughout this paper.

The minimal triplet in (9) illustrates the three-way deictic spatial distinction:

(9) (SBL, elicited)

a. *ɣuro'o-teè* *daiŋ=rè*
big-INT hill=PROX
'Este cerro es grande.'
'This hill is big.'

b. *ɣuro'o teè* *daiŋ=re*
big-INT hill=DIST
'Ese cerro es grande.'
'That hill is big.'

c. *ɣuro'o-teè* *daiŋ=ki*
big-INT hill=INVIS
'Ese cerro (que no podemos ver) es grande.'
'That hill (not visible) is big.'

The proximal *=rè* is used for referents very close to the speaker, distal *=re* for things that are further away but (generally) still visible, and *=ki* for objects that are not visible to the speaker at the time of speaking. There is no evidence that the position of the addressee relative to the referent has any importance. In the following minimal pair, *=re* is used to refer to mountains that are actually not visible to the speaker, but where the interlocutors are able to look towards the direction of the mountain. When asked how the meaning would change if *=ki* were substituted, SMG explained that *=ki* would be used if you could not see the mountains at all, even to point in their direction, such as when you are inside and talking about them.

(10) (SMG, elicited)

a. *kɛd* *ri-ɛnlo-di-uŋ* *ð-dain=re*
NEG HAB-see-NEG-1P PL-mountain=DIST
'We can't see those mountains.'

b. *kɛd* *ri-ɛnlo-di-uŋ* *ð-dain=ki*
NEG HAB-see-NEG-1P PL-mountain=INVIS
'We can't see those mountains.'

The choice of Spanish demonstrative adjective (*ese* versus *aquel*) in elicitation did not have an effect on the *=re* versus *=ki* distinction; *=re* seemed to be the default translation for both and *=ki* would generally only be used when a situation was set up where it was made clear that the referent was not visible. According to SMB, the use of *=ki* implies that you have seen the referent before, or that it is known to the interlocutors (as in (11)); thus, the extension into a past tense marker is not surprising.

(11) (SBL, elicited)
kun *libr=ki*
where book=INVIS
'¿Donde está el libro (que estaba aquí antes)?'
'Where is the book (that was here before)?'

Finally, although *=ki* is usually used when talking about a referent in the past, it would not be used if the referent happened to be visible at the time of speaking, as in the example in (12), in which the mountain in question was understood to be visible.

(12) (SBL, elicited)
na *ɣu-zunè-è* *daiŋ=rè*
I COMPL-climb-1S hill=PROX
'Yo subí este cerro.'
'I climbed this hill.'

The fourth determiner *=kaŋ* may indicate an intermediate distance between *=rè* and *=re*, but consultants do not generally seem to classify it as belonging to the same set. Its primary function seems to be an indicative one, and it is probable that its sphere of spatial reference sometimes overlaps the aforementioned pair. Two consultants, SBL and SMG, said that *=kaŋ* is used when pointing to the referent (as in (13a)), but pointing is not a requirement for its use, and there are a few tokens of *=kaŋ* where it seems to function as a definite article in the same way as the other determiners (as in (14)). However, it is possible in some of these examples that the consultant was imagining a situation where he or she was pointing out a referent or picking out a specific entity from a set.

(13) (SMG, elicited)
a. (This sentence was produced as a translation for the English prompt. When asked if the meaning would change if *=re* were substituted for *=kaŋ,* SMG said that the *=kaŋ* sentence would be used if you were pointing at the pot.)

tʃi *ri-ab* *nis* *ɣiɛ* *ri-dʒa-iŋ* *nis* *ɣɛs=kaŋ*
when HAB-fall water stone HAB-fill-3S.INAN water pot=MED
'When it rains the pot fills with water.'

b. *tʃi* *ri-ab* *nis* *ɣiɛ* *ri-dʒa-iŋ* *nis* *ɣɛs=re*
when HAB-fall water stone HAB-fill-3S.INAN water pot=DIST
'When it rains the pot fills with water.'

(14) (MMC, elicited)
bitʃi'in na gusɛɛŋ=kaŋ
small COP girl=MED
'La niña es chiquita.'
'The girl is small.'

The instances of *=kaŋ* in the texts are consistent with a possible indexical interpretation; see (25) below.

Although the texts provided a large number of tokens of the determiners used in connected speech, there were no examples that clearly showed the expected spatial interpretation. There simply were no situations where objects at different distances from the speaker were contrasted, and in fact there were cases in which different determiners were used to describe the same object in the same spatial position (see section 6). Targeted elicitation, with a clearly defined context and often with real-world referents, easily provided the evidence to establish the basic spatial sense of these determiners. As Mithun (2001:38) found, "[e]licitation can be crucial for filling in paradigms, for securing forms we can predict to exist." Once the spatial sense was clearly established, however, I found that I was lacking the same types of clear examples for the temporal sense—those that can be provided by elicited minimal pairs and carefully constructed prompts. Thus, although the original analysis of the texts had led to the temporal interpretation, a new round of elicitation was needed to refine it—an approach to fieldwork advocated by Chelliah (2001:152), "in which text collection and elicitation are interwoven in a finely tuned, constantly modulated way."

5. THE TEMPORAL SENSE: TEXTS AND ELICITATION. The most commonly used determiner in the TdVZ texts, which included folktales, local legends and personal narratives, was the distal *=ki*, which is often glossed as indicating 'past tense'. As mentioned above, the extension from spatial into temporal deixis is a natural one; an entity that the speaker has knowledge of, but that is not currently visible, is generally one that he or she saw sometime in the past. Often, the use of *=ki* coincides with the completive aspect, but it also appears in sentences lacking a verb or where the verb is in another aspect. In these cases, the choice of determiner in the NP provides additional temporal cues about an event or state. There are many instances of these types of temporal cues in the texts, but elicited minimal pairs are the most effective way to illustrate the phenomenon. In the following minimal pair, volunteered by LGL as an example, the stative copula on its own gives no indication of when the woman was seen by the speaker, but the choice of determiner tells the listener whether she is present at the time of utterance, or was there at an earlier time.

(15) (LGL, example given by consultant to explain meaning of *=ki*)
a. *sru-teè na gunaa=ki*
nice-INT COP woman=INVIS
'Que bonita es la mujer. (i.e., de ayer, poquito pasado)'
'That woman is beautiful. (i.e., was here yesterday, or a little bit in the past)'

b. *sru-teè* *na* *gunaa=re*
nice-INT COP woman=DIST
'Que bonita es esa mujer. (i.e., que esta aquí)'
'That woman is beautiful. (i.e., who is present now)'

If the consultant's interpretation were put aside, it could be argued that the determiners in (15) are purely spatial and the past tense interpretation of (15a) is merely inferred from the context. The following pair provides examples of an unambiguous temporal interpretation. As noted above, *=re* is sometimes used for referents that are not strictly visible at the time of utterance, but which are still somehow more accessible than *=ki* referents. *=ki* implies remoteness of the referent, and that remoteness extends to the event or situation in which the referent is involved. In (16), *=re* refers to a man who has just left the scene, while the *=ki* sentence implies that the situation (the man being 'here') took place further in the past. The verb in both sentences is in the completive aspect, so only the choice of determiners differentiates between recent and more remote past events.

(16) (LGL, elicited)

a. *tu* *bɛni=re* *gu-zuwa'a* *rè*
who person=DIST COMPL-be here
'¿Quien fue este hombre que estuvo aquí?'
'Who was that man who was (just) here?'

b. *tu* *bɛni=ki* *gu-zuwa'a* *rè*
who person=INVIS COMPL-be here
'¿Quien fue este hombre que estuvo aquí ?'
'Who was that man who was here? (i.e., more in the past; it could mean he was here yesterday.)'

The determiner clitics are often used with abstract referents that by their nature are not visible and could not have any spatial orientation relative to the speaker. In these examples, the choice of determiner correlates with a temporal distinction.

(17) (LGL, elicited)

a. *kwɛnt=rè* *gu-in* *ni* *dʒooban-tè-uŋ*
story=PROX POT-do that sad-INT-1P
'El cuento nos va a hacer muy triste.'
'The story is going to make us really sad.'

b. *kwɛnt=ki* *ba-in* *ni* *dʒooban-tè-uŋ*
story=INVIS COMPL-do that sad-INT-1P
'El cuento hizo que estamos muy triste. (i.e., el cuento ya esta contado)'
'The story made us really sad. (i.e., it has already been told)'

In the texts, the clitic *=ki* was always used to refer to referents temporally located in the past, but targeted elicitation yielded further possibilities. Just as the choice of determiner can distinguish between recent and more remote past (as in (16)), it may also differentiate between near and more distant future. The former is illustrated by the use of *=re* in (17a), while the latter is illustrated by (18).

(18) (SMG, elicited)

ʒi	*tʃin-a'*	*rupte*	*ð-gunaa*	*bak=ki*
tomorrow	POT.go.see-1S	the.two	PL.woman	Tlacolula=INVIS

'Mañana voy a ver / visitar las dos mujeres de Tlacolula.'

'Tomorrow I'm going to see the two women from Tlacolula.'

The determiners are often used with nouns that express units of time. For the following example, ALB explains that *=rè* is used when referring to the year that has just past (*el año recién pasado*) and *=ki* denotes a more distant past year; it is less specific but could be two, three or more years ago.

(19) (ALB, elicited)

a.

iz	*gu-dɛɛð=rè*	*ba-tʃiib-tè-uŋ*	*ʒi'in*	*bell-a'*
year	COMPL-pass=PROX	COMPL-scare-INT-1P	son	sister-1S

'El año recién pasado asustamos al hijo de mi hermana.'

'This past year we really scared my sister's son.'

b.

iz	*gu-dɛɛð=ki*	*ba-tʃiib-tè-uŋ*	*ʒi'in*	*bell-a'*
year	COMPL-pass=INVIS	COMPL-scare-INT-1P	son	sister-1S

'El año pasado asustamos al hijo de mi hermana.'

'That year we really scared my sister's son.'

Some other common temporal expressions that make use of the determiners are *na'a=rè* 'right now' ('now' + PROX) and *tʃi=ki* 'at that time' ('when' + DIST).

Elicitation is also a very important tool in creating examples of ungrammatical sentences, and in finding out what is not possible in a language. In this study it was used to test whether the TdVZ determiner clitics followed any of the same patterns as temporal markers on the NP in other languages. Nordlinger & Sadler (2004) identify two types of tense/aspect/mood (TAM) marking on NP constituents: "Propositional Nominal TAM, whereby the nominal carries TAM information relevant to the whole proposition, and Independent Nominal TAM, in which the TAM information encoded on the nominal is relevant only to the NP on which it is marked" (p.1). Propositional Nominal TAM, as described in Nordlinger & Sadler (2004), is a phenomenon where the clause-level TAM is expressed on the NP, either by case or other morphology that works in conjunction with the verb to express clausal TAM, or by morphemes that appear only on the noun but have a non-local interpretation. This is not the case in TdVZ, where the determiners have scope only over the NP and do not show any type of agreement behavior with the verbal aspect prefixes. The canonical type of Independent Nominal TAM is that which occurs in Guaraní and other

languages of the Americas, in which a noun *x* may be marked with a past tense suffix that would generate a meaning such as 'the former *x*', or in the case of a possessed NP, 'my former *x*', as illustrated by (20) ((3) and (6) in Nordlinger & Sadler (2004)).

(20) (Gregores & Suárez 1967:127)

a. *h-óga-kwé*
his-house-PST
'his former house'

b. *o-va-ta* *che-róga-kue-pe*
3-move-FUT 1SG-house-PST-in
'He will move into my former house.'

I have not found this type of nominal TAM marking to occur in TdVZ; in elicitation, this kind of information was usually expressed with a relative clause, as in (21) and (22).

(21) (SBL, elicited)

na'a	*laadaŋ*	*zuɣwa'a-daŋ*	*ni*	*ɣuk*	*liiz-a'*
now	they	HAB.be/stay-3P	REL	COMPL.be	house-1S

'Ahora ellos viven en mi casa (de antes).'
'Now they live in my former house.'

(22) (SMG, elicited)

ni	*guk*	*tʃiel*	*bizaan-a'*	*ka-zuuz-aŋ*
REL	COMPL.be	husband	sister-1S	PROG-get.drunk-3S

'El ex-esposo de mi hermana esta emborrachándose.'
'My sister's ex-husband is getting drunk.'

For the following elicited examples, the prompt was making reference to a previously discussed (hypothetical) store that had closed down and was no longer on the site. SMG first produced the sentence in (23a), but when asked if *=re* could be replaced by *=ki*, he gave the sentence in (23b) and said that it would be a more specific translation, and that (23a) had been a direct translation of the Spanish and did not express the fact that the store was no longer there.

(23) (SMG, elicited)

a.

bɛn	*gunaa=re*	*guk*	*ʃteenj-aŋ*	*diɛnd=re*
person	woman=DIST	COMPL.be	POSS-3S	store=DIST

'Esa mujer era la dueña de esa tienda.'
'That woman was the owner of that (former) store.'

b. *bɛn* *gunaa=re* *guk* *ʃteenj-aŋ* *diɛnd* *ni* *kɛtiru=ki*
person woman=DIST COMPL.be POSS-3S store REL NEG.be-INVIS
'Esa mujer era la dueña de esa tienda.'
'That woman was the owner of that store that isn't there.'

According to Klein (1994), tense concerns the relation between 'topic time' (TT), defined as the time for which a claim or assertion is made, and the 'time of utterance' (TU). In this view, the 'situation time' (TSit) is not relevant to tense; the time of the situation being described may extend beyond the boundaries of TT in either direction. It does not make any assertion about the relationship between the TSit and TU for any property of the referent. This is in contrast to the canonical type of Independent Nominal TAM discussed in Nordlinger & Sadler (2004), where the addition of a temporal maker to a noun makes an assertion about the TSit of a noun having some property or being in someone's possession. Using the Klein (1994) framework, in its temporal sense I analyze *=ki* as a tense marker establishing a topic time that is somewhat remote from the time of utterance, or more precisely, as placing the nominal referent in a remote TT relative to the TU. In the following example, *=ki* refers to some workers—the people who do work for one of the characters in the story. The use of the distal determiner, in addition to having anaphoric discourse reference, places the workers in a remote TT, but does not make the assertion that these people are no longer working for the character at TU or any other time. We only know that they were habitually working for him at the TT being referenced. To make the assertion that they formerly worked for him, or had already completed their work for him, the verb *ruin* 'do' in the relative clause would have to be in the completive aspect.

(24) (LGL, text: *Pobreza*)
bɛni *ni* *ru-in* *dziinʲ-aŋ=ki* *aʂ* *ba-in-daŋ* *gwa*
person REL HAB-do work-3S=INVIS and COMPL-do-3P COMPL.go

gwɛɛ-dɛɛ-dan-iŋ
COMPL.go-give-3P-3S.OBJ
'Sus trabajadores lo fueron a dar …'
'The people who do work for him (his workers) then went to give it...'

Although the determiner may only have scope over the NP, the interpretations of the minimal pairs given by consultants for many of the above examples do show that the determiner can influence the interpretation of the whole proposition by giving more specific temporal cues than those that can be expressed by the verbal aspect alone. Because the minimal pairs were elicited and were not from naturally-occurring speech, they tell us the possible uses of the determiners to make fine spatial and temporal distinctions, but not whether these uses occur commonly in regular speech. Analysis of the texts, however, does show frequent discourse-based usage of the determiners that makes use of their temporal properties.

6. DISCOURSE FUNCTIONS: TEXTS. As mentioned above, there were consultants who gave the Spanish determiners as translations for the TdVZ determiners, and some glossed *=ki* as *tiempo pasado* 'past tense'. Others had difficulty coming up with a specific Spanish word to gloss any of the determiners, identifying them as discourse markers, or as VC explained, "something used to help the discourse." This type of usage can really only be understood through the study of natural speech, for reasons noted by Mithun:

> An obvious value of the documentation of natural connected speech is that it permits us to notice distinctions and patterns that we might not know enough to elicit, and that might not even be sufficiently accessible to the consciousness of speakers to be volunteered or retrievable under direct questioning. (Mithun 2001:45)

In this section I present some of the discourse functions of the determiners that have been used in the texts. Because the texts were of certain genres (e.g., folktale, local or personal history), probably only a small set of the possible uses of the determiners was represented. Examination of other types of naturally occurring discourse, such as conversations and procedural texts, will most likely reveal further uses of the determiners.

In many of the texts collected for this project, the story is told mainly by a narrator, but there are many instances of quoted speech in which one of the characters is either speaking to another or thinking out loud. Along with the shifts between narrated and quoted text, there is almost always a corresponding shift in the determiners that are used, with the distal *=ki* being the main determiner used by the narrator. When a character from the story is speaking, a second TU is established, and the TT of the quoted utterance falls within that new TU, so the proximate determiners are used. The following example illustrates the shift between narrated and quoted speech. The character in the story has died, and is at the point in his journey to the otherworld in which he has to cross a river. The boatman has given him a choice of two canoes in which to cross, one that costs two *reales* (a type of coin) and another that costs four. The use of *=ki* versus *=re* or *=kaŋ* clearly marks off the narrator's speech from that of the two characters involved in the dialogue.

(25) (LGL, text: *No Sabemos Para Quien Trabajamos*)

per	*laaŋ*	*kɛtị̊*	*nɛɛ-di-aŋ*	*ʧi-ɛn*	*lɛ'ɛ*	*ni*	*sakt*	*aph*	*rel=ki*
but	he	neg	say-neg-3s	pot.go-3s	in	rel	cost	four	real=invis

'Pero él no quiere ir adentro del que cuesta cuatro reales.'

'But he doesn't want to go in the one that costs four *reales*.'

aṣ	*ne-ŋ*	*ʧa'-a*	*lɛ'in*	*ni*	*sak*	*tjop*	***reel=kaŋ***
and	say-3S	POT.go-1S	in	REL	cost	two	real=MED

rɛ'ɛbi-aŋ *bɛni=**ki.***
HAB.say-3S person=INVIS
'Y dijo, "Quiero ir adentro en la que cuesta dos reales", le dijo a la persona.'
'And he says, "I want to go in the one that costs two *reales*", he says to the person.'

ɣubi'n-teè *ɣuk-u* *lo:* *hɛṣliu* *rɛ'ɛbi-aŋ* *laaŋ*
cheap-INT COMPL.be-2S on earth HAB.say-3S 3S
'"Fuiste muy avaro en la tierra", le dijo a él.'
'"You were really cheap on Earth", he said to him.'

na'a *na* *par* *kizo'o* *ni* *sak* *tap* *reel=**re***
now COP for POT.pay REL cost four real=DIST
'"Ahora tienes que pagar la que cuesta cuatro reales."'
'"Now you have to pay for the one that costs four *reales*."'

The frugal man decides to save the four *reales* and swim across the river. Once he gets to the other side he continues walking, then reaches his destination and meets God. When God asks him why he has arrived wet, the man gives the following explanation, using the determiner *=re* to refer to the canoe even though it is no longer within sight.

(26) (LGL, text: *No Sabemos Para Quien Trabajamos*)
aṣ *ne-ŋ* *es* *ke* *kɛth* *ri-ka'az-di-a'* *nihi'iʒ-a* *kanoa*
and say-3S it's that NEG HAB-want-NEG-1S pay-1S canoe
'Y le dijo, "Es que no quise pagar la canoa'
'...and he said, "it's that I didn't want to pay for the canoe'

ni *sak* *taph* *reel=re* *rɛ'ɛbi-aŋ* *laaŋ*
REL cost four real=DIST HAB.say-3S 3S
'que cuesta cuatro reales", le dijo a él.'
'that cost four reales", he said to him.'

Both *=ki* and *=re* have anaphoric discourse reference, but the former is used when the 'remote' narrator is relating past events, and the latter when the ongoing story is told from the perspective of a character involved in the action, establishing a secondary TU-TT relationship contrasting with the matrix TU-TT interface.

In (27), the use of *=ki* in the last clause, referring to a previously mentioned bottle, identifies that portion of the text (*kon ṣkal ɣubi tar=ki*) as a comment by the narrator, following the quoted thoughts of the character, which are related in first person. Without *=ki*, it would have been ambiguous as to whether that clause was part of the first person narrative or something expressed by the 'remote' narrator.

(27) (LGL, text: *La Bruja Saco a Erasto*)

ne-ŋ *ßel* *ɣa-baiŋ* *d-ßeni=rè* *tʃi-a-ɣa-in-daŋ*
say-3S if POT-wake.up PL-person=PROX go-already-POT-do-3P
'y dijo, "si se despiertan estas gentes (sic) van a'
'he says, "If these people wake up they are going to'

tʃi-atiʒlɛ'ɛ-daŋ *o* *hustis* *ʃi-nde'e* *ka-junj-a'* *rè* *nadʒab*
POT-go.tell-3P to police what-this PROG-do-1S here bad
'ir a avisar a la justicia que estoy haciendo acá, haciendo malo,'
'go tell the police what I am doing here, doing bad things,'

kon *ʂkal* *ɣu-bi* *tar=ki*
with how COMPL-make.noise bottle=INVIS
'como sonó el bote."'
'because of how the bottle made noise."'

With most NPs that are not inherently definite, the determiners are used consistently when a specific, previously introduced referent is mentioned in the discourse. For example, in the story "El Tecolote y la Culebra," the owl, snake and man are indexed with a determiner throughout the story (see (4)), and specific inanimate referents such as the canoe in (25) similarly fail to appear without a determiner. As mentioned above, possessed nouns generally do not take a determiner. In the following example, both referents have been previously introduced, but only the non-possessed form is marked with a determiner.

(28) (LGL, text: *Zopilote*)

beni *gunaa=ki* *ra-beè-teè-ŋ* *ʒigab* *ni* *ka-jun*
person woman=INVIS HAB-be.happy-INT-3S idea REL PROG-do

tʃiel-aŋ
husband-3S
'Esa mujer (está) muy contenta de lo que está pensando su esposo.'
'The woman was happy with the idea that her husband had.'

However, there are a few cases where inalienably possessed NPs are marked with a determiner. In the following example, which is from a story about two brothers, the narrator alternates between the possessed form without a determiner and the possessed form followed by *=ki*. I assume this serves a sort of tracking function, to clarify that the same brother who was mentioned in the previous sentence is being referenced.

(29) (LGL, text: *Pobreza*)

ri-kaaz-aŋ *hiakbeen* *ʃa* *ɣuk* *te* *zu-gwa'a* ***bɛts-aŋ***
HAB-want-3S know how COMPL.be that HAB.sit brother-3S
'Quiere saber como fue que su hermano vive'
'He wants to know how his brother lives'

zɛkᵢ̯ *zin-teè* *nɛɛz* *gu-dili-aŋ* *per* ***bɛts-aŋ=ki*** *kɛt*
that.way a.lot-INT road COMPL-search-3S but brother-3S=INVIS NEG
'así. Buscó muchas formas (de saber) pero su hermano no'
'in that manner. He searched for a way (to find out) but his brother didn't'

ri-ʒe-di-aŋ *ʃa* *ba-anʲ-aŋ* *sru* *per* *tʃi* *ba-ʒiad-aŋ*
HAB-tell-NEG-3S how COMPL-do-3S well but when COMPL-find-3s
'le contó como ha cambiado. Pero cuando encontró la forma,'
'tell him how he did well. But then he found the way,'

aᵢ̯ *ba-suuz-aŋ* ***bɛts-aŋ*** *lo* *gal-rizuuz=ki* *aᵢ̯*
and COMPL-get.drunk-3S brother-3S on NMLZ-drunkenness=INVIS and
'luego enborrachó su hermano. En la borrachera'
'and he got his brother drunk. In his drunkenness'

gu-diʒlɛ'ɛ ***bɛts-aŋ=ki*** *ʃa* *ɣuktin* *a-ri-bɛɛz-aŋ* *sru-teè*
compl-tell brother=INVIS how COMPL-be already-HAB-sit-3S well-INT
'le contó su hermano como fue de que se entra bien.'
'the brother told how it was that he started doing well.'

In the narrative *Historia de Pancho Villa*, LLV makes frequent reference to *benᵢ̯ daiŋ* 'people from the mountains'. Although she sometimes uses a determiner with this phrase, more often she does not. It is likely that it is not always the same group of people being referenced, but rather different individuals from the mountains. However, there are some cases where the context indicates that it is likely the same individuals being mentioned, and the determiner fails to appear. In (30), the people from the mountain are first introduced into the discourse without a determiner, as expected; they are referenced a second time with *=ki*, as expected; then they are mentioned a third time, but without a determiner where one would be expected.

(30) (LLV, text: *Historia de Pancho Villa*)

ɣa-zuin ***bɛn*** ***daiŋ*** *lo* *zie* ***ßɛn*** ***daiŋ=ki*** *aᵢ̯*
POT-arrive person mountain on COMPL-go person mountain=invis and
'Llegaron las personas de la montaña. Se fueron las personas de la montaña y luego'
'The people from the mountains arrived. The people from the mountain went and later'

bɛ'ɛt-an	*aʂ*	*ze*	*da nelį gutier*	*ze-ŋ*	*lua'ą*
COMPL-come-3P	and	COMPL-go	Don Manuel Gutiérrez	COMPL-go-3S	Oaxaca

'vinieron y luego se fue Don Manuel Gutiérrez a Oaxaca.'
'they came and Don Manuel Gutiérrez went to Oaxaca.'

aʂ	*tʃi*	*ba-za-an*	*lua'ą*	*aʂ*	*bi-zuin*	***bɛnį***	***daiŋ***
and	when	COMPL-return-3S	Oaxaca	and	COMPL-arrive	person	mountain

'Y cuando regresó de Oaxaca, luego llegaron las personas de la montaña.'
'And when he returned from Oaxaca, then the people from the mountains arrived.'

More research on naturally occurring speech is needed to explain the optionality of the determiners in some circumstances. It may be a matter of stylistic differences between speakers or genres.

7. CONCLUSIONS. Demonstratives are traditionally described as indexing definiteness and deictic spatial relations, and most Zapotec (and other) grammars give only a brief description with an overview of the type of spatial distinctions they make. However, although the spatial sense may be the primary one from which the other uses are derived, the examination of naturally occurring discourse shows that the 'secondary' uses of this set of determiners are actually the most common, especially for the distal determiner *=ki*. This detailed study of one small portion of the language's grammar shows the importance of using different types of data in language documentation. Targeted elicitation and discussions with native speaker consultants were used to bring out the spatial reference of the determiners, and also to rule out certain types of Independent Nominal TAM. In addition, elicitation brought out other possible discourse uses that had not occurred in the texts, such as the emphatic effect obtained by using a determiner twice for a single referent. The texts illustrated the discourse-level uses of the determiners, which primarily made use of their temporal properties; these uses would have been difficult to discover through elicitation alone. Starting the work on the language with the study of natural connected speech such as narratives also helped to shape this investigation. Rather than the linguists coming up with a list of words and phrases in a contact language for the consultants to translate, in this case the consultants provided the Zapotec words and sentences that we were to work with, and then defined them on their own terms. Although the glosses provided by the consultants at first seemed to be inconsistent, these native speaker intuitions helped to guide a more thorough investigation that tested each possibility given. In the end, each consultant had provided an important piece of the puzzle in describing these multifunctional determiners.

Appendix: Texts cited

Lorenzo Gutiérrez Lopez, 2004. Te damm kon te bell / El tecolote y la culebra.
Lorenzo Gutiérrez Lopez, 2005. La bruja sacó a Erasto.
Lorenzo Gutiérrez Lopez, 2005. Xa guk te bia'an gallzi loo gezhliu / Como fue que la pobreza vive en el mundo.
Lorenzo Gutiérrez Lopez, 2005. Kedih Nadondyon Tupar Ri'enyun Zïin / No sabemos para quien trabajamos.
Lorenzo Gutiérrez Lopez, 2007. Xa guk te laam gul bets / Como fue que el zopilote se llama 'gul bets'.
Lucia Lazaro Vasquez, 2005. Historia de Pancho Villa (translated by Viridiana Chavez).

References

Angulo, Jaime de. 1926. Tone patterns and verb forms in a dialect of Zapotek. *Language* 2(4). 238-250.

Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic fieldwork. In Paul Newman & Martha Ratliff (eds.). 152-165.

de Angulo, Jaime (see Angulo)

Givón, Talmy. 2001. *Syntax,* vol. 1. Amsterdam & Philadelphia: John Benjamins.

Gordon, Raymond G., Jr. (ed.). 2005. *Ethnologue: Languages of the world,* 15th edn. Dallas: SIL International. http://www.ethnologue.com (10 May, 2008.)

Gregores, Emma & Jorge Suárez. 1967. *A description of colloquial Guaraní*. The Hague: Mouton.

Instituto Nacional de Estadística Geografía e Informática (INEGI). 2008. Conteo 2005. http://www.inegi.gob.mx/est/contenidos/espanol/sistemas/conteo2005 (10 May, 2008.)

Klein. Wolfgang. 1994. *Time in language.* London & New York: Routledge.

Lillehaugen, Brook Danielle. 2006. *Expressing location in Tlacolula Valley Zapotec*. Los Angeles, CA: University of California dissertation.

Lowes, Gwendolyn & Ausencia Lopez Cruz. 2007. Potential aspect and the role of tone in two variants of Zapotec. In Eloina Miyares Bermúdez & Leonel Ruiz Miyares (eds.), *Linguistics in the twenty-first century*, 135-144. Cambridge: Cambridge Scholars Press.

Mithun, Marianne. 2001. Who shapes the record: The speaker and the linguist. In Paul Newman & Martha Ratliff (eds.). 34-54.

Munro, Pamela & Felipe H. Lopez. 1999. *San Lucas Quiavini Zapotec dictionary*. Los Angeles: UCLA Chicano Studies Research Center.

Newman, Paul & Martha Ratliff (eds.)*, Linguistic fieldwork.* Cambridge: Cambridge University Press.

Nordlinger, Rachel & Louisa Sadler. 2004. Nominal tense in crosslinguistic perspective. *Language* 80(4). 776-806

Pickett, Velma B. & Cheryl Black. 1998. *Gramatica popular del Zapoteco del Istmo*. Juchitán: Instituto Lingüistico de Verano A.C. y Centro de Investigación y Desarrollo Binniza.

Smith-Stark, Thomas C. 2003. Algunas isoglosas Zapotecas. Paper presented at the III Coloquio International de Lingüística. México, D.F.

Donna Fenton
donna@lmntl.com

7

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 147-171
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/6789

Middles and reflexives in Yucatec Maya: Trusting speaker intuition[1]

Israel Martínez Corripio[a] and Ricardo Maldonado[b]

[a]*Escuela Nacional de Antropología e Historia*
[b]*Universidad Nacional Autónoma de México/Universidad Autónoma de Querétaro*

In this paper we provide a characterization of the middle construction in YM, and show that the apparently unpredictable distribution of middle voice in YM corresponds to a neatly identified, and quite limited, system of *absolute events*, i.e., events in which no energy is expended (Langacker 1987). This strategy is not exploited by other related Mayan languages, which tend to encode all absolute events as simple intransitive verbs. The semantic coherence of middle voice in YM is only discernible by combining analysis of narrative texts and direct elicitation with attention to speaker intuition in a variety of situational contrastive contexts guided by cognitive principles which are known to determine the behavior of middle voice systems in other languages.

1. INTRODUCTION. Middle voice in Yucatec Maya (YM) is well known for being a small portion of the syntax that presents aberrant and unpredictable behavior. Some uses resemble what is found in middle voice in other languages of the world, namely that the event remains in the emotional/physical/mental sphere of the subject. Yet unlike other languages with a middle system, neither the semantic nor the syntactic restrictions seem to follow any predictable pattern (Ayres & Pfeiler 1997; Bohnemeyer 2004; Briceño 2004; Bricker 1981, 1998). The goal of this paper is twofold. First, we provide a characterization of the middle construction in YM, and show that the apparently aberrant behavior of middle voice in YM corresponds to a neatly identified, and quite limited, system of *absolute events*, i.e., events in which no energy is expended (Langacker 1987). Such a restricted semantic area is rather uncommon among languages having a middle system. We will suggest that the middle system in YM corresponds to one specific strategy of construing absolute events from transitive active clauses, while other absolute construals follow different patterns. This strategy is not exploited by other related Mayan languages, which tend to encode all absolute events as simple intransitive verbs.

Our second goal for this paper is to show that the subtle nature of middle voice in YM requires not only data derived from direct elicitation and narrative texts, but also attention

[1] We want to thank Doña Consuelo Chan for her invaluable help and generous time. Thanks are also due to Professor Fidencio Briceño at the INAH, Yucatán, who patiently listened to our multiple attempts to handle middle voice in Yucatec Maya. Finally, special thanks are due to an anonymous reviewer whose comments helped us tie up important loose ends in our analysis.

to a variety of situational contrastive contexts guided by cognitive principles which are known to determine the behavior of middle voice systems in other languages. Our fieldwork methodology is presented in section 1.3 below. Because middles, reflexives, and root intransitive verbs contrast in subtle ways in YM, fine-grained semantic and pragmatic information must be taken into account to provide a systematic representation of the middle system. Our data collection began with direct elicitation and the analysis of oral narrative, but these—whether alone or considered together—were not sufficient to fully illuminate the behavior of the YM middle system. As our analysis grew, we found it necessary to invent ways to investigate speaker intuition as well.

1.1. THE RESEARCH QUESTION. Let us first identify the middle construction. Middle voice is defined as a system of constructions whose main property is to portray events remaining in the dominion of the subject (Kemmer 1993, 1994). In middle voice constructions, the action is performed with special reference to the subject: the subject is, in most cases, an experiencer undergoing the action expressed by the verb. Since the action remains in the dominion of the subject, the similarity between middle and reflexive constructions is considerable. Yet in YM, neatly contrastive reflexive-middle pairs are common, as in (1):

(1) a. Reflexive
t-u-ts'ak-(*aj*)-*ø* *u=**ba***
PERF.TRS-A3S-heal-COMP.TRS-B3S P3S=REF[2]
'He cured himself' (ConChan: 05/2005)

b. Middle
*ø-**ts'áak**-ø-i*(*j*)
PERF.INTR-heal.MID-COMP.INTR-B3S
'He got better (he became cured)' (ConChan: 05/2005)

Both middles and reflexives develop from active transitive constructions. Reflexive constructions are formed by the possessive marker inflected for person and number, plus a reflexive marker *-ba,* which imposes a coreferential reading on a (root or derived) transitive verb, like *ts'ak* 'heal' in (1a). Middle voice constructions have a CV́VC pattern, obtained by lengthening the vowel of a transitive active CVC verb stem (e.g., *ts'ak*) and imposing a high tone on the first vocalic segment (*ts'áak*) (Ayres & Pfeiler 1997, Briceño 2004, Bricker 1981, Bohnemeyer 2004), as in (1b). In the most transparent cases, the reflexive/middle contrast coincides with well-known cross-linguistic patterns: in reflexive constructions, the

[2] In this paper we use the following abbreviations; A='A series (ergative)'; AP='antipassive'; APL='applicative'; B='B series (absolutive)'; CAUS='causative'; CL='classifier'; COMP='completive'; DEM='demonstrative'; DUR='durative'; FEM='feminine'; HAB='habitual'; IMM.FUT='immediate future'; INCOMP='incompletive'; INTR='intransitive'; MASC='masculine'; MID='middle', NEG='negative'; OBL='oblique'; P='possessive'; PASV='passive'; PE='emphatic pronoun'; PERF='perfective'; PL='plural'; REF='reflexive'; S='singular'; SUB='subjunctive'; TOP='topic'; TRS='transitive'; = = bound morpheme; 1='1st person'; 2='2nd person'; 3='3rd person'.

subject acts volitionally and with control on the self, inducing some change. In contrast, middles highlight the subject's affectedness as a consequence of some change-of-state the subject undergoes but does not volitionally control. Thus the reflexive construction in (1a) designates the subject acting deliberately to improve his health (e.g., by taking medicine), while in (1b) the subject's health simply improves with no particular effort.

While the middle/reflexive contrast is clear, the behavior of middle constructions in the system has resisted a clear and systematic representation (Ayres & Pfeiler 1997; Briceño 2004; Bricker 1981, 1988). First, the middle seems to be restricted to a reduced number of verb classes. They can only be derived from transitive CVC verbs (Bricker 1981, 1998; Briceño 2004; Bohnemeyer 2004). Moreover, not all transitive situations can take the middle marker. For example, self-care situations can take the reflexive marker *-ba* as in (2a) and (3a) but not middle marking as in (2b) and (3b):

(2) a. *t-u-chal-*(*aj*)*-ø* *u=ba*
PERF.TRS-A3S-refresh-COMP.TRS-B3S P3S=REF
'He refreshed himself' (ConChan: 05/20e05)

b. **ø-cháal-ø-i*(*j*)
PERF.INTR-refresh.MID-COMP.INTR-B3S
Intended reading 'He refreshed'

(3) a. *t-u-k'os-*(*aj*)*-ø* *u=ba*
PERF.TRS-A3S-cut.with.scissors-COMP.TRS-B3S P3S=REF
'He cut with scissors (his hair)'(ConChan: 05/2005)

b. **ø-k'óos-ø-i*(*j*)
PERF.INTR-cut.with.scissors.MID-COMP.INTR-B3S
Intended reading: 'He cut with scissors (his hair)'

Middle constructions can be formed with verbs of motion, but verbs of change-of-position like *kul* 'sit' (4a) can only take reflexive marking, as in (4b), not middle marking, as shown in (4c):

(4) a. *k-u-kul-tal*
HAB-A3S-sit-INCOMP.INTR
'He sits down' (ConChan: 05/2005)

b. *k-u-kul-kint-ik-ø* *u=ba*
HAB-A3S-sit-CAUS-INCOMP.TRS-B3S P3S=REF
'He sits himself down' (ConChan: 05/2005)

c. **k-u-kúul-tal*
HAB-A3S-sit.MID-INCOMP.INTR
Intended reading: 'He sits himself down'

Of particular interest is the fact that middles and reflexives in YM contrast in ways that have resisted a clear explanation. As (5) shows, the closing of a door is an event that can be encoded by either the reflexive or the middle:

(5) a. *t-u-k'al-*(*aj*)*-ø* *u=ba* *le* *joonaj-o'*
PERF.TRS-A3S-close-COMP.TRS-B3S P3S=REF DEM door-DEM[3]
'The door shut (by itself)' (ConChan: 05/2005)

b. *ø-k'áal-ø-ø* (*l*)*e* *joonaj-o'*
PERF.INTR-close.MID-COMP.INTR-B3S DEM door-DEM
'The door shut' (ConChan: 05/2005)

Moreover, in cases where either the reflexive or the middle construction can be used, the reflexive implies a reading of unexpectedness, as in (6a). Crucially, the middle construction in (6b) would be chosen to depict natural occurrences (e.g., a firecracker bursting in the town plaza during festivities):

(6) a. *t-u-xik-*(*aj*)*-ø* *u=ba* *le* *boolador-o'*
PERF.TRS-A3S-burst-COMP.TRS-B3S P3S=REF DEM firecracker-DEM
'The firecracker burst' (by itself) (ConChan: 05/2005)

b. *ø-xíik-ø-ø* (*l*)*e* *boolador-o'*
PERF.INTR-burst.MID-COMP.INTR-B3S DEM firecracker-DEM
'The firecracker burst' (ConChan: 05/2005)

As far as we know, no previous analysis has acknowledged this contrast. We hope to be able to account for the variety of ways in which middles and reflexives interact in YM. We will also attempt to delineate the semantic area occupied by the middle construction and show how it interacts with other absolute construals in the system.

1.2. THE YM MIDDLE VOICE SYSTEM IN CROSS-LINGUISTIC CONTEXT. We find that middle voice constructions in YM constitute a well-organized and coherent system. In YM *reflexives* codify complex single-participant *energetic* events; i.e., events in which the subject exhibits high degree of control of his actions, while *middles* in Yucatec Maya simplify complex transitive events into *absolute* simplex events. As proposed by Langacker (1987), the contrast between *absolute* and *energetic* events corresponds to the degree of energy profiled in the event. In absolute events the energy being used to carry an action is neutral and is not relevant to the mental representation of the event, regardless of whether the event involves different degrees of energy input. In contrast, the profile of the energy implied or required by an energentic event is crucial for the proper mental representation of the event.

[3] Demonstratives in YM take an initial and a terminal particle. For sake of simplicity we limit ourselves to marking both as DEM.

This contrast is evident in French: absolute events in the composite past tense take the auxiliary *être* 'be'. In verbs like *aller* 'go', *venir* 'come', and *arriver* 'arrive', no speed or manner of locomotion is put forward, and *être* is the proper auxiliary. Energetic verbs, however, take the auxiliary *avoir* 'have', and present the opposite configuration: they profile speed and manner of locomotion, as in *courir* 'run', *nager* 'swim', *voler* 'fly', and so on. A similar contrast can be seen in a variety of languages. Langacker hypothesizes that the three morphological classes of Cupeño, as reported by Hill (1969), correspond to the same contrast. Verbs with the suffix *-ine* are volitional and active, while those marked with *-yaxe* tend to have the opposite properties. As opposed to these polar suffixes, verbs with zero-marking are energy-neutral; these include most verbs for states of mind (e.g., 'be angry'), natural bodily processes (e.g., 'see'), and the behavior of plants, animals, inanimate subjects, and the weather (e.g., 'bloom', 'rain'). In Cupeño, zero-marked forms refer predominantly to processes that run according to the normal, expected course of events as they are conceived of in the physical domain or in the cultural sphere. Actions that involve considerable input of energy, like hunting and skinning, may still count as absolutes because they are part of Cupeño everyday culture.

The behavior of the Spanish clitic *se*, traditionally analyzed as reflexive, has been analyzed in terms of a similar contrast (Maldonado 1988, 1999, 2009). Verbs of motion in Spanish (e.g., *subir* 'go up', *bajar* 'go down', *salir* 'go out', *ir* 'go') may be used as plain intransitives if they simply describe the path followed by the subject, without consideration of manner, speed, or degree of energy input. When speed and energy are profiled, however, the clitic *se* must be used; the event is seen as rapid sudden and energetic. Thus, if one goes up a mountain, one might say *subí la montaña* 'I went up the mountain', using a plain intransitive form. On the other hand, if a rat suddenly comes into a room, a frightened speaker who jumps on the sofa would likely use the reflexive to say *me subí al sofa* 'I got (i.e., jumped, leaped) on the sofa', indicating the sudden quality of the movement. In Spanish the clitic *se* marks energetic events and follows general tendencies of middle marking for languages that have such a system. Middle marking in YM does not follow those tendencies. We provide an account below for YM's exceptional middle marking.

In YM, we find that middle voice constitutes a strategy to derive absolute events from energetic ones. Absolute middles contrast with reflexives precisely in that the former do not profile energy, but they derive simplex events from complex ones. The contrast between complex and simplex events is determined by the degree of control of the subject over the depicted action. On the one hand, complex events involve an agent acting on him or herself volitionally, much in the same way that transitive constructions involve an agent acting upon a theme. Simplex events, on the other hand, are restricted to changes of state undergone by a thematic subject. In simplex events, the subject is, in most cases, an experiencer undergoing some change of state. If these assumptions are supported by the data, we can explain both the unpredictable distribution of middles in YM and the surprising similarities between middles and root intransitive verbs. We claim that, as opposed to the close relationship between middle and reflexive voices found in many languages, YM middles and reflexives do not overlap. In fact, in certain contexts, the two constructions are mutually exclusive. The constructions we examine are not simply syntactic ensembles, but rather respond to specific semantic and pragmatic conditions implied by the complex representation of such constructions.

From a typological perspective, Kemmer (1993, 1994) has identified a set of situations where the middle tends to occur across languages. Table 1, adapted from Kemmer (1993), is a sample of situations in which we may expect to find middle marking.

The semantic areas where the middle tends to occur can be characterized in more general terms: motion (translational motion, non translational motion, and change in body posture), change of mental state (cognition, emotive speech, and emotion), and spontaneous events and self-directed actions (direct and indirect body care). Of all these categories, the one that fluctuates between middle and reflexive marking across languages is that of self-directed actions; the other three tend to receive middle marking. As we show, the middle system in YM presents restrictions not commonly found in middle systems of other languages. Although it is well known that self-care situations vary between reflexive and middle marking from language to language, typical cases of translational and non-translational motion, in which we expected middles, did not follow the expected pattern, while spontaneous events and other cases derived from transitive verbs did take middle marking. In some texts, an intransitive or a middle construction contrasted with a reflexive construction with no evident difference in context. Only after checking with our consultants about the possible interpretation of the contrast were we able to come up with the correct analysis for each context. As we shall show in the conclusion section, the analysis of the middle system in YM was only possible through a detailed consideration of the interaction between middles, intransitives, and reflexive constructions in pragmatic context.

TABLE 1. Basic middle voice situations

Grooming or body care	
Latin	*lavor* 'wash'
Indonesian	*berdandan* 'get dressed'
Non translational motion	
Kanuri	*tàn-t-în* 'stretch one's body'
Latin	*revertor* 'turn'
Change in body posture	
Indonesian	*berlutut* 'kneel-down'
Guugu Yimidhirr	*dagaadhi* 'sit down'
Translational motion	
Pangwa	*i-nu-xa* 'climb up'
Guugu Yimidhirr	*madhaadhi* 'climb up'
French	*s'en aller* 'go away'
Indirect middle	
Turkish	*edin* 'acquire'
Cl. Greek	*ktasthai* 'acquire for oneself'
Emotion middle	
Guugu Yimidhirr	*dumibaadhi* 'get a shock or fright'
Mohave	*mat iθa:v* 'be angry'
Hungarian	*bánkod* 'grieve, mourn'
Emotive speech actions	
Latin	*queror* 'complain'
Cl. Greek	*olophyresthai* 'lament'
Turkish	*dövün* 'lament'
Cognition middle	
Indonesian	*berpikir* 'be cogitating'
Pangwa	*-isala* 'think over, consider'
Spontaneous events	
Indonesian	*berhenti* 'come to a stop'
French	*s'evatiouir* 'vanish'
Hungarian	*keletkez-* 'originate, occur'

1.3. METHODOLOGY. We first noticed the behavior of the reflexive marker *-ba* in 2004 during our fieldwork in Ekmul (Yucatán). We found that several types of intransitive verbs (positionals, de-adjectivals, and those with inanimate subjects) would take an unexpected reflexive marker in some contexts. We then checked for dialectal differences among speakers of Tixkokob and Holcá (Yucatán), where we also found *-ba* occurring in more verbs of the same classes as those we had seen in Ekmul.

Because we had been working primarily with consultants and had not yet investigated texts, we suspected that perhaps these constructions were restricted to informal spoken language. We then examined written narrative texts, where we found similar unexpected reflexive marking on intransitive verbs. At this point we began to suspect errors in our data, so we double-checked our data in 2005 with Doña Consuelo Chan from Holcá. Our data thus confirmed, we then began to formulate hypotheses about the syntactic and pragmatic environments that might trigger the unexpected use of *-ba.*

In 2007 we returned to the field to test our hypotheses by presenting speakers with contrasting contexts for intransitives, middles, and reflexives. For example, we found in text data that 'kneel' could be expressed with both an intransitive and a reflexive construction, and we inferred that the latter implied fast motion. Based on the examples we had for 'kneel', we created hypothetical situations involving other actions (e.g., moving out of the path of a projectile) and invented examples to discover which constructions would be grammatical in particular contexts. Speaker intuitions about subtle semantic and contextual differences proved invaluable to us. In some cases, speakers provided hypothetical scenarios (such as practical jokes) in which our invented examples could be considered grammatical. These subtly contrastive contexts allowed us to formulate the analysis presented here about the behavior of middles in YM, and we provide many of our ungrammatical invented examples in this paper to illustrate the analysis.

The remainder of this paper is organized as follows: section 2 provides some basic YM structural information relevant for the analysis. In Section 3 we show how reflexive, plain intransitives, and middles contrast, and we offer an account of the distribution of their semantic space. We propose that the distinction between middles and reflexives is determined by the contrast between absolute and energetic events. Given this account, we also offer an explanation for reflexive constructions that designate unexpected events. In section 4 we conclude and suggest some further considerations about the absolute/energetic contrast as pertaining to the function of middle voice marking.

2. BASIC STRUCTURAL INFORMATION ABOUT YM. Yucatec Maya is an ergative language with a split system in intransitive verbs. The split is determined by aspect: in the imperfective aspect the alignment is accusative, with the intransitive subject marked like the subject of a transitive clause. As examples (7a, b) show, the subject pronoun takes the (A series) first person ergative pronoun *in-*.

(7) Imperfective
  a. *táan* ***in****-wen-el*
    DUR A1S-sleep-INCOMP.INTR
    'I am sleeping' (ConChan: 05/2005)

b. *táan* ***in****-ween-s-ik-****ech***
DUR A1S-sleep-CAUS-INCOMP.TRS-B2S
'I am putting you to sleep' (ConChan: 05/2005)

On the other hand, perfective aspect causes ergative alignment. The subject in (8) takes the (B series) second person absolutive marker *-ech*, which corresponds to the object marker of a transitive clause:

(8) Perfective
(*j*)*=wen-ø-ech*
PERF.INTR=sleep-COMP.INTR-B2S
'you slept' (ConChan: 05/2005)

As will be evident from the vast majority of our examples, middle voice is usually marked in the perfective aspect.

In YM there is a three-way contrast between transitive, reflexive, and middle constructions. (9a) and (10a) show the transitive base from which both reflexives and middles are formed. The reflexive takes the possessive prefix linked to the reflexive marker *-ba* as in (9b) and (10b), while the middle construction is encoded by lengthening the nuclear vowel of the verb with a high tone as in (9c) and (10c):

(9) a. *leken* ***u-ts'am-(i)k-ø*** (*l*)*e* *luuch-o'* *le* *máak-o'*
when A3S=submerge-COMP.TRS-B3S DEM bowl-DEM DEM person-DEM

k-u-jóok-ol *uy*[4] *óom* *ja'-o'*
HAB-A3S=go.out-INCOMP.INTR P3S bubble water-DEM
'when that person sinks the bowl in the water, the bubble comes up'
(ConChan: 11/2007)

b. ***t-u-ts'am-(aj)-ø*** ***u=ba***
PERF.TRS-A3S-submerge-COMP.TRS-B3S P3S=REF

ichil *ja'* *le* *paal-o'* *chéen* *báaxal*
inside water DEM child-DEM only play

k-u-meen-t-ik-ø
HAB-A3S-do-APL-INCOMP.TRS-B3S
'the child submerges in the water he goes in just to play'
(ConChan: 11/2007)

[4] The [y] or [w] is an epenthetic glide which is introduced in coda position to avoid an undesired vocalic sequence.

c. ***ø-ts'áam-ø-ø*** (*l*)*e* *nook'-o'*
PERF.INTR-submerge.MID-COMP.INTR-B3S DEM clothing-DEM

yáan *u-p'o'-ik- ø* *sáama*(*l*)
OBL A3S-wash-INCOMP.TRS.B3S tomorrow
'the clothes submerged in the water, she has to wash them tomorrow' (ConChan: 11/2007)

(10) a. ***yáan*** ***a-sats'-*(*i*)*k-ø*** (*l*)*e* *suum-o'*
OBL A2S-pull-INCOMP.TRS-B3S DEM rope-DEM

ka *a-k'ax-(i)k-ø* (*l*)*e* *ba'alche-o'*
then A2S-tie-INCOMP.TRS-B3S DEM animal-DEM
'you have to pull the rope to then tie the animal' (ConChan: 11/2007)

b. *leken* *uy-áaj-al* (*l*)*e* *miis-o'*
when A3S-wake.up-INCOMP.INTR DEM cat-DEM

t-u-sats'-(aj)-ø ***u=ba***
PERF.TRS-A3S-stretch-COMP.TRS-B3S P3S=REF
'when the cat wakes up it stretches out' (ConChan: 11/2007)

c. *le* *nook'-o'* *ø-sáats'-ø-i*(*j*)
DEM garment-DEM PERF.INTR-strech.MID-COMP.INTR-B3S
'the clothes stretched' (ConChan: 05/2005)

The contrast is clear. Transitives involve an agentive subject acting on a thematic object. Reflexives resemble transitives in having two distinguishable participants (Kemmer 1993), but the agent acts on a coreferential theme. Middles, however, are quite different. These involve an event where the thematic subject simply undergoes some change. What is not as clear is the precise point at which the two coreferential participants become a single inseparable participant undergoing change. In the next section, we do, however, discuss the general restrictions on coreferentiality in YM that help locate the 'breaking point' between reflexive and middle constructions.

3. REFLEXIVES AND MIDDLES IN YUCATEC MAYA. There is a general restriction in YM that only one entity may be coreferent with the subject. Thus YM has no indirect reflexives like those commonly found in some Indo-European languages, in which the two coreferent elements are in a part/whole relationship. This is the case for Spanish: consider *Adrián se lavó la cara* 'Adrián washed his face', in which *se* 'REF' and *cara* 'face' are coreferent with *Adrián*. In YM a named body part precludes the reflexive marker *-ba*, as in (11a). Conversely the presence of the reflexive *-ba* (11b) precludes the use of *ook* 'foot'. The ungrammaticality of example (11c) shows that coreferential parts and wholes are mutually exclusive:

(11) a. ***yáan*** ***a-ts'ak-(i)k-ø*** *aw* *ook* *k-a-pat-(i)k-ø*
OBL A2S-cure-INCOMP.TRS-B3S P2S foot HAB-A2S-wait-INCOMP.TRS-B3S

a-bíin *páak*
A2S-go weed.out
'you have to wait and heal your foot to go and weed out (the field)'
(ConChan:11/2007)

b. *pwes* *le* *máak-o'* *mina'an* *taak'in* *ka* *tujuuna*
well DEM person-DEM nothing money then only

t-u-ts'ak-(aj)-ø ***u=ba***
COMP.TRS-A3S-heal-COMP.TRS-B3S P3S=REF
'well, that person does not have money thus he cure himself'
(ConChan:11/2007)

c. **k-in-ts'ak-ik-ø* *in=ba* *inw* *ook*
HAB-A1S-heal-INCOMP.TRS-B3S P1S=REF P1S foot
Intended reading: 'I cure myself on the foot'

Although it is possible to find cases where both middles and reflexives seem to overlap, a review of the whole system shows that in fact the area covered by middles is quite restricted. It should not be surprising, then, to find self-affecting actions encoded only by reflexives. To the extent that the subject actually acts on him or herself, the reflexive is the natural option, as shown in (12a). The possibility of using the middle construction in self-affecting actions is banned in YM, as can be seen from (12b):

(12) a. *chéen* *k-u-laj-ik-ø* *yaana(l)* *máak-o'ob*
only HAB-A3S-slap-INCOMP.TRS-B3S another person-PL

leti'-e' *ma'* ***t-u-laj-(i)k-ø*** *u=ba*
PE3S-DEM NEG DUR-A3S-slap-INCOMP.TRS-B3S P3S=REF
'he only slaps other people, he doesn't slap himself' (ConChan: 11/2007)

b. **ø-láaj-ø-i(j)*
PERF.INTR-slap.MID-COMP.INTR-B3S
Intended reading: 'he slapped himself'

The same restriction takes place in grooming verbs. Only the reflexive is available to encode self-care actions, as can be seen from the consistent use of the reflexive marker *-ba* in (13a), (14a), and (15a) and the ungrammaticality of their middle counterparts in (13b), (14b), and (15b):

(13) a. *le máak-o' bul k'iin t-u-meyaj-ø*
DEM person-DEM all day DUR-A3S-work-INCOMP.INTR

leken k-u-xu(l)-(i)k-ø t(i')-u taanaj-e'
when HAB-A3S-finish-INCOMP.TRS-B3S in-P3S house-DEM

k-u-chal-(i)k-ø u=ba
HAB-A3S-rinse-INCOMP.TRS-B3S P3S=REF
'that person is working all day, when he finishes he washes at home'
(ConChan: 11/2007)

b. **ø-cháal-ø-i(j)*
PERF.INTR-rinse.MID-COMP.INTR-B3S
Intended reading: 'he rinsed himself'

(14) a. *le x-ch'úupal-o' t-uy-il-aj-ø u taatá*
DEM FEM-girl-DEM PERF.TRS-A3S-see-COMP.TRS-B3S P3S father

t-u-k'os-(i)k-ø u=ba
PERF.TRS-A3S-cut.with.scissors-COMP.TRS-B3S P3S=REF
'the girl saw that her father is cutting his hair' (ConChan: 11/2007)

b. **ø-k'óos-ø -i(j)*
PERF.INTR-cut.hair.MID-COMP.INTR-B3S
Intended reading: 'he cut his hair'

(15) a. *uláak' k'iin-e' t-uy=il-aj-ø*
another day-TOP PERF.TRS-A3S=see-COMP.TRS-B3S

táan u-ts'ik-ik-ø u=ba u ts'uulil-o'
DUR A3S-shave-COMP.TRS-B3S P3S=REF P3S boss-DEM
'another day he saw that his boss was shaving' (Mono: 518)

b. **ø-ts'íik-ø-i'(j)*
PERF.INTR-shave.MID-COMP.INTR-B3S
'he shaved himself'

The way grooming actions are encoded may shed some light on the restricted properties of middle voice in YM. Grooming actions are commonly marked by middles in languages that have such a system, however in YM they are only expressed as reflexives, as shown in (13a), (14a), and (15a). Interestingly, grooming verbs in YM can also take antipassive marking, as in (16) and (17). Like reflexives, antipassives are operations imposed on the object rather than on the subject. In the antipassive construction, the object becomes non-specific and thus the event becomes generic:

(16) *ba'ax* *t-u-meen-t-aj-ø* *le* *x-ch'úupal*
what PERF.TRS-A3S-do-APL-COMP.TRS-B3S DEM FEM-girl

tu'ux *j-bíin-ø-ø-o'* *chéen* ***ø-k'oos-naj-i(j)***
where PERF.INTR-go-COMP-B3S-DEM only PERF.INTR-cut.AP-COMP.INTR-B3S
'What did the girl do where she went? She only cut with scissors (hair)' (ConChan: 11/2007)

(17) *káaj* *u-bat-ik-ø* *ø-laaj-naj-i(j)*
start A3S-fight-INCOMP.TRS-B3S PERF.INTR-slap.AP-COMP.INTR-B3S

le *x-ch'úupal-o'* *káal-chaj-i(j)* *ka* ***laaj-naj-i(j)***
DEM FEM-girl-DEM drunk-COMP.INTR-B3S then slap.AP-COMP.INTR-B3S
'She started to fight with him and slapped (him). That girl got drunk and then she slapped (him).' (ConChan: 11/2007)

Middles constitute operations imposed on the subject as a theme, and thus the possibility of encoding grooming actions is simply blocked. The (invented) middle forms **láaj* and **k'óos* are not possible in YM. Verbs of motion behave similarly.

Cross-linguistically, situations involving the subject's change-of-position are often marked as middles. This is to be expected, because self-induced action is only observable in the change-of-position itself; the energy expended is not externally evident, since the energy necessary to impose the change of position is applied internally. We do not 'see' our legs acting on a body part when sitting down or standing up. Interestingly, as with verbs of grooming, change-of-position verbs in YM take reflexive marking, not middle marking. This suggests that the YM middle system covers a restricted semantic area within the range of possibilities normally covered by middle voice in most languages having such a system.

Of particular interest is the fact that in YM there are two ways to encode change of position. The reflexive construction contrasts with root intransitive verbs to highlight different views of the same event. Note the contrast between (18), where the intransitive *kul* 'sit down' depicts a routine action, and (19), where the reflexive shows a high degree of control. In (18) the subject Hees simply sits down to see the king:

(18) *ka* ***kul-laj-ø*** *Hees* (*uy*)-*il*-(*aj*)
then sit-COMP.INTR-B3S Hees A3S-see-COMP.TRS

bix *uy* *ok'ol-ø* *le* *réey-o'*
how A3S cry-INCOMP.INTRS DEM king-DEM
'then Hees sat to see how the king would cry' (Hees: 262)

In (19), drawn from another story, the subject's change of position is not simply a routine action, but is instead a strong reaction to noise that awakens the main character. He sits up quickly to see what is going on:

(19) *ka* *chil-laj-ø* *wenel...* *ka* *t-uy-u'u-b-(aj)-ø*
then lay-COMP.INTR-B3S sleep then PERF.TRS-A3S=hear-PASV-COMP-B3S

u *yúuch-ul* *áakan* *te'* *tu'ux* *yan-o'...*
A3S happen-INCOMP.INTR shout there where be-DEM

k-u-kul-kint-ik-ø ***u=ba***
HAB-A3S-sit.down-CAUS-INCOMP.TRS-B3S P3S=REF
'then he laid down to sleep and then he heard a scream right where he was and he got up right away' (Flojo: 110)

The use of the intransitive verb corresponds to background information, i.e., normal occurrences with no special relevance for the specific context of the event. In contrast, the reflexive construction can be used to foreground events, and especially to profile an energetic event. In these cases a causative marker is also used, which renders the construction transitive. In (19), the character's action of sitting up in reaction to a noise is of special relevance to the story. The contrast is corroborated in (20) by the intransitive verb *xol* 'kneel' in (20a) and the reflexive version with *-ba* in (20b):

(20) a. *ka* ***xol-(l)aj-ø*** *le* *ko'olel-o'*
then kneel-COMP.INTR-B3S DEM woman-DEM

ka *jo'op'* *u-payalchi'*
then start A3S-pray
'then the woman knelt down and started to pray' (ConChan: 11/2007)

b. *ka* *t-uy-il-aj-ø* *jun=túul* *nuxi'* *ooch*
then PERF.TRS-A3S-see-COMP.TRS-B3S one=CL old fox

ka *t-u-xol-kint-aj-ø* *u=ba*
then PERF.TRS-A3S-kneel-CAUS-COMP.TRS-B3S P3S=REF
'then he saw the old fox and he knelt fast'(Flojo: 122)

In (20b) the action of kneeling is done quickly, in order to avoid being seen by the fox; here the action is marked with reflexive morphology. A routine instance of kneeling, e.g., in church as in (20a), simply requires an intransitive verb.

Verbs deriving from adjectives contrast with reflexives in the same way that root intransitives do. While the reflexive construction depicts an energetic event, the de-adjectival intransitive verb encodes a neutral absolute event. In (21a) the absolute intransitive encodes a natural and effortless change of body temperature; in contrast, the reflexive form

in (21b) encodes a controlled situation where the subject purposefully lowers his or her temperature before going out into the open air.[5]

(21) a. *ø-síis-chaj-ø* *tumen* *ø-chil-laj-ø*
PERF.INTR-cold-COMP.INTR-B3S because PERF.INTR-lying.down-COMP.INTR-B3S

lu'um
ground
'he got cold because he was lying down on the floor' (ConChan: 05/2005)

b. *leken* *áaj-ak-e'* ***yáan*** ***a-síis-kunt-(aj)-ø*** ***a=ba***
when stand-SUB-DEM OBL A2S-cold-CAUS-COMP.TRS-B3S P2S=REF

utia'al *a-bin* *meyaj*
for A2S-go work
'when you wake up, cool off before you go to work'(ConChan: 11/2007)

Control is even more evident in the case of getting sick. In the reflexive construction in (22b) the subject volitionally makes himself sick in order to avoid going to work. The absolute intransitive in (22a) is a spontaneous uncontrolled event:

(22) a. *le* *ko'olel-o'* ***ø-k'oja'an-chaj-i(j)***
DEM woman-DEM PERF.INTR-get.sick-COMP.INTR-B3S

k'abeet *u-bíin-ø* *jo'*
necessary A3S-go-INCOMP.INTR Mérida
'The woman got sick, she needs to go to Mérida'(ConChan: 11/2007)

b. *t-u-k'oja'an-kunt-(aj)-ø* *u=ba*
PERF.TRS-A3S-sick-CAUS-COMP.TRS-B3S P3S=REF

ti'olal *ma'* *u* *bin-ø* *meyaj*
in.order.to NEG A3S go-INCOMP.INTR work
'He got sick to avoid going to work' (ConChan: 05/2005)

Let us now observe the behavior of middle constructions in YM. Recall that the middle construction marks absolute events, i.e., changes of state that are not controlled by the thematic subject, while the reflexive construction reflects an energetic version of the same basic event. One can simply get better, as in the middle construction (23a), or volitionally take action to improve one's physical state, as the reflexive examples in (23b) and (23c) illustrate:

[5] In Mexico there is the belief that a sudden change of body temperature may cause a serious illness (eye paralysis, pneumonia, and so forth).

(23) a. *bix* *u* *beel* *aw* *aal*
how P3S way P2S daughter

ø-ts'áak-ø-i(j) *táan* *u-meyaj-ø*
PERF.INTR-cure.MID-COMP.INTR-B3S DUR A3S-work-INCOMP.INTR
'how is your daughter? she is healed she is already working'
(ConChan:11/2007)

b. *sáansamal* *t-u-ts'ak-(aj)-ø* *u=ba* *yéetel* *xíiw*
everyday PERF.TRS-A3S-cure-COMP.TRS-B3S P3S=REF with herbs
'everyday he cures himself with herbs' (ConChan: 05/2005)

c. *le* *j-meen-o'* ***t-u-ts'ak-ø*** ***u=ba***
DEM MASC-healer-DEM DUR-A3S-cure-COMP.TRS P3S=REF

tujuuna *chéen* *leti'-e'* *uy-o(j)-(i)k* *bixi'*
only only PE3S-DEM A3S-know-INCOMP.TRS how
'the healer cured himself and only he knows how' (ConChan:11/2007)

Likewise, with verbs of translational motion, the middle construction is appropriate if the subject asserts no energy or control. However, if the change of location is the consequence of a controlled action, only the reflexive can be used. When the child goes into the water to get a coin the reflexive marker *-ba* is used (24b), but the middle construction is adequate for cases where the subject simply sinks without control or energy, as would be the case for drowning, as in (24a):

(24) a. *ba'ax* *k-u-n* *j-táal-ø-ø*
what HAB-A3S-IMM.FUT PERF.INTR-come-COMP.INTR-B3S

ø-búul-ø-i(j)
PERF.INTR-sink.MID-COMP.INTR-B3S
'how is he going to come? he already drowned' (Nazario: 92)

b. *k-u-pul-(i)k-ø* *(l)e* *taak'in-e'* *ichil* *le*
HAB-A3S-throw-INCOMP.INTR-B3S DEM money-DEM inside DEM

ja'-o' *ka* *t-u-bul-(aj)-ø* *u=ba* *(l)e*
water-DEM then PERF.TRS-A3S-sink-COMP.TRS-B3S P3S=REF DEM

paal-a'
child-DEM
'the child throws the coin into the water and he dives in to get it'
(ConChan: 05/2005)

The variety of cases seen so far suggests a general tendency in YM to construe events in an absolute manner in which the expended energy is not profiled. The event may involve no energy at all, as in the case for drowning; in other cases the energy driving the event may simply be part of the basic semantics. In YM current events tend to be expressed in absolute, neutral terms. Absolute events may be encoded in at least three ways: with root intransitive verbs of motion or change of position, with verbs deriving from adjectives, and with middle constructions. We propose that the middle is a construction used specifically to derive absolute events from transitive energetic verbs. These three strategies contrast drastically with the reflexive construction, which encodes energetic events involving only one participant.[6] While it is true that the reflexive can be used for self-directed actions with a split-represented participant as in (13a), (14a), and (15a), the construction has expanded in YM to include a range of single-participant events in which the action is under the control of the subject.

In the previous examples the subject is consistently volitional and human, but the following examples show that non-human subjects may also be used with the reflexive construction. Note in (25) that the extinguishing of the fire may be conceptualized in two ways, reflected in the choice of either middle (25a) or reflexive (25b) marking:

(25) a. *káaj-tal* *cháak-e'* *ka* ***ø-túup-ø-ø***
start.INCOMP.INTR rain-TOP then PERF.INTR-go.off.MID-COMP.INTR-B3S

le *k'áak-o'*
DEM fire-DEM
'it started to rain, then the candle went out' (ConChan: 05/2005)

b. *le* *paal-o* *ka* *j-óok-ø-ø* *ichil*
DEM child-DEM then PERF.INTR-went.in-COMP.INTR-B3S inside

le *áaktun-o'* *ka* *t-u-**tup**-(aj)-ø* *u=**ba***
DEM cave-DEM then PERF.TRS-A3S-go.off-COMP.TRS-B3S P3S=REF

tujuuna *u* *kíib*
alone P3S candle
'the child entered the cave and the candle went out by itself'
(ConChan: 05/2005)

The contrast is easily explained by our analysis. Since the middle encodes absolute construals, when the flame simply dies as a natural phenomenon the middle marker is the natural choice. However, when the flame dies out against natural expectations, the reflexive marker is used, as in (25b). This is not an isolated example. The contrast is fully established in the language and can be identified given the right context. Another example showing the

[6] This generalization of course does not apply to reciprocal constructions where *–ba* involves more than one participant. We thank an anonymous reviewer for calling our attention to this matter.

systematicity of the pattern is found in (26), where an identical middle/reflexive contrast takes place. If, for example, one holds a party at which the guests eat a large number of tortillas, one fully expects to run out of corn. Under those circumstances, the lack of corn is a natural expected event which is properly expressed by the middle marker, as in (26a); however if one is hungry and suddenly notices that there is no corn, the contradicting event must be expressed by the reflexive construction, as in (26b):

(26) a. *mina'an-e'* *waaj-o'* *tumen* ***ø-xúup-ø-ø***
none-TOP tortilla-DEM because PERF.INTR-finish.MID-COMP.INTR-B3S

(l)e *ixi'im-o'*
DEM corn-DEM
'there are no tortillas because we are out of corn'(ConChan: 11/2007)

b. *ø-líik'-ø-ø* *le* *máak-o'* *ka*
PERF.INTR-stand.up-COMP.INTR-B3S DEM person-DEM then

t-uy-il-(*aj*)*-ø* *t-u-**xup***-(*aj*)*-ø*
PERF.TRS-A3S-see-COMP.TRS-B3S PERF.TRS-A3S-finish-COMP.TRS-B3S

*u=**ba*** *le* *ixi'im-o'*
P3S=REF DEM corn-DEM
'that person stood up and noticed that the corn was gone'
(ConChan: 11/2007)

One last example illustrates the productivity of this pattern. The spontaneous neutral event gets middle marking, as in (27a), and the unexpected event gets reflexive marking, as in (27b). In both cases the subject is inanimate, and yet only in the reflexive construction can we get a reading of unexpectedness:

(27) a. *ma'* *inw=óok-s-*(*i*)*k-ø* *ja'* *le* *in*
NEG A1S=go.out-CAUS-INCOMP.TRS-B3S water DEM P1S

luuch-a' ***ø-jóol-ø-i*(*j*)**
bowl-DEM PERF.INTR-hole.MID-COMP.INTR-B3S
'I cannot get any water my bowl has a hole in it' (ConChan: 05/2005)

b. *ts'o'ok* *in-náak-al* *inw* *ool* *in-wach'-(i)k-ø*
finish A1S-fed.up-INCOMP.INTR P1S soul A1S-untie-INCOMP.TRS-B3S

(*l*)*e* *kisin* *suum-a'* ***t-u-ch'ot-m(aj)-ø-ø***
DEM demon rope-DEM PERF.TRS-A3S-twist-COMP.TRS-B3S

u=ba ***tujuun-a'***
P3S=REF only-DEM

' I am fed up of untangling this damned rope, it got all twisted by itself.'
(ConChan:05/2005)

We may wonder why the reflexive construction is employed to express counter-expectations. The answer is simple. As shown above, the subject of a reflexive construction is always a volitional human. If this human subject requirement is not met, this fact *in itself* contradicts the natural configuration of a reflexive construction. We propose that in cases of nonhuman subjects, a reflexive *-ba* construction expressing accidental or unexpected events will occur. The nonhuman subject *-ba* construction is a clear example of iconicity in the sense of Haiman (1980, 1983), in which the structural organization of the construction reflects the semantic-pragmatic content of its use. The simplex event is encoded by middle marking while the more complex type of construal is marked for reflexive. The reflexive encodes situation in which a human volitionally acts on self while the middle marks natural spontaneous events. The complexity of the construal matches the complexity of the marking form. In the case of the unexpected construction the complexity of the construal is even more evident. At the semantic-syntactic level, the non-humanness of the subject goes against the fundamental requirement of a reflexive construction. At the pragmatic level, the construction is used to express situations running counter to expectation, in which inanimate elements are participants and take action. To the extent that non-humans are occupying a position reserved for humans in control, the unexpected construction can be said to be iconic and highly motivated.

There are two mutually complementary explanations for the function of middles. The middle construction, together with other constructions in YM, depicts absolute events, which are energy-neutral, ordinary, and part of everyday life in Mayan culture. The middle contrasts both pragmatically and syntactically with the energetic reflexive construction. YM derives absolute events with middle voice from active transitive and obviously energetic roots. Middles are not just absolute; they are absolutes *contrasting* with energetic events. In other words, the main function of middles is to *contrast* with either an active transitive or with a reflexive construction.

4. CONCLUSIONS. This paper provides a systematic account of middle constructions in YM. In order to explain the behavior of middles, we looked at the behavior of the whole set of related intransitive constructions in the verbal system. The similarities among constructions involving only one participant allowed us to observe that there is a more general contrast between absolute and energetic constructions. Energetic events correspond to transitive active constructions as well as reflexive constructions. In both cases the subject is human and has volitional control over his actions. In the absence of a volitional human sub-

ject the reflexive construction is still energetic, but a new meaning of 'counter-expectation' is obtained. This sense of counter-expectation comes from the fact that the subject is non-animate (not energetic) and thus cannot have volitional control over the action depicted by the verb. On the other hand, absolute constructions include several types of intransitive one-participant events derived from different sources: root intransitive verbs, adjectives, and active transitive verbs. Reflexives and middles develop intransitive events from the same transitive stem; the reflexive preserves the energetic configuration of the verb, while the middle derives a verb whose energy is not profiled.

We propose that the three intransitive constructions (root intransitive, de-adjectival, and motion verbs) and found in YM form a more general superclass of absolutes. Supporting this argument is the fact that there are about ten intransitive verbs that exhibit a long vowel with high tone. While syntactically these verbs are unquestionably intransitive, they are shaped phonologically like middles (see Table 2). This mixed representation suggests an important overlap between intransitives and middles.

TABLE 2. Intransitive verbs with a long vowel with high tone

jóok' 'go out'	*kóoj* 'arrive'
síij 'born'	*tóop'* 'bloom/emerge'
xúuk 'arrive suddenly'	*kíim* 'die'
líik' 'get up'	*lúub* 'fall'
p'áat 'remain'	*síij* 'be born'

Interestingly these verbs do not derive from transitive active verbs. To make them transitive, a causative marker *-s* must be inserted, in the same way that transitive verbs are derived from root intransitives. Moreover, to make a reflexive construction the verb must also take the causative marker. In other words, the verb must first become transitive in order to *then* become reflexive. A plain intransitive is shown in (28a). Example (28b) is its reflexive *-ba* counterpart with the causative *-s* marker. From the ungrammaticality of (28c) we know that the causative marker *-s* is obligatory. Similarly (28d) is ungrammatical because it lacks a high toned long vowel:

(28) a. *ka* *jo'op' le* *cháak-o' le* *wakax-o'* **(*j*)*-lúub-ø-i*(*j*)**
then start DEM rain-DEM DEM cow-DEM PERF.INTR-fall-COMP.INTR-B3S
'Then it started to rain and the cow fell down' (ConChan: 11/2007)

b. ***t-u-lúu*(*b*)*-s-*(*aj*)*-ø*** ***u*=*ba*** *tuyo'olal*
PERF.TRS-A3S-fall-CAUS-COMP.TRS-B3S P3S=REF in.order.to

ma' *u-ko'och-ol* *tumen* *wakax*
NEG A3S-run.over.PASV-INCOMP.INTR by cow
'he let himself fall to avoid being run over by the bull' (ConChan: 11/2007)

c. **Waan-e'* *t-u-**lúub**-(i)k-ø* *u=ba*
Juan-TOP DUR-A3S-fall-INCOMP.TRS-B3S P3S=REF
Intended reading: 'Juan lets himself fall (so as not to fall down from the tree)'

d. **Waan-e'* *t-u-**lub**-(i)k-ø* *u=ba*
Juan-TOP DUR-A3S-fall-INCOMP.TRS-B3S P3S=REF
Intended reading: 'Juan lets himself fall (so as not to fall down from the tree)'

The behavior of these verbs follows that of absolute root intransitive verbs with a short vowel, as in (29). Note that the causative marker *-s* is required to form the transitive in (29b), and that the same marker is required to make the reflexive construction in (29c):

(29) a. *t-in-**wen**-el*
DUR-A1S=sleep-INCOMP.INTR
'I am sleeping' (ConChan: 05/2005)

b. *t-in-**ween**[7]**-s**-ik-ø* *le* *paal-o'*
DUR-A1S-sleep-CAUS-INCOMP.TRS-B3S DEM child-DEM
'I am putting the baby to sleep' (ConChan: 05/2005)

c. *t-in-**ween-s**-ik-ø* *in=ba*
DUR-A1S-sleep-CAUS-INCOMP.TRS-B3S P1S=REF
'I am going to sleep' (ConChan: 05/2005)

The historical emergence of intransitive verbs with a long vowel and high tone requires independent investigation. For the current study, it is sufficient to say that the existence of these verbs suggests not only a close familiarity between middles and intransitives in YM, but also considerable overlap between the two. They in fact constitute a bigger class of absolutes where only the change of state is profiled leaving aside any potential inducing force.

Figures 1-5 represent a Cognitive Grammar (Langacker 1987, 2000) interpretation of the YM system. On the right are the set of energetic constructions. Subject and object are equated with trajector (Tr) and landmark (Lm) as the first and second most prominent participants in the event. The circle represents a participant, the double arrow stands for the transmission of energy from subject to object and the squiggly arrow represents the change of state impose by the action. Figure 1 shows the basic representation of a transitive active clause. The reflexive construction in Figure 2 shows the same properties with a dotted line indicating coferentiality connecting subject and object. In Figure 3, all the properties of the reflexive construction are present, but the counter-expectation construction comes from the fact that the subject is a theme instead of an agent. As for the absolute subsystem, we

[7] For many of these intransitive verbs the short vowel lengthens but maintains low tone when they take the causative marker. This phenomenon is not related to the behavior of the middle construction.

give two representations: Figure 4 stands for a variety of intransitive constructions (root intransitive, de-adjectival and motion verbs), and Figure 5 represents the middle construction. Here the subject undergoes the affectedness imposed by the verb as depicted by the arrow returning to the circle. The subject is a thematic participant unable to exert control of his or her actions.

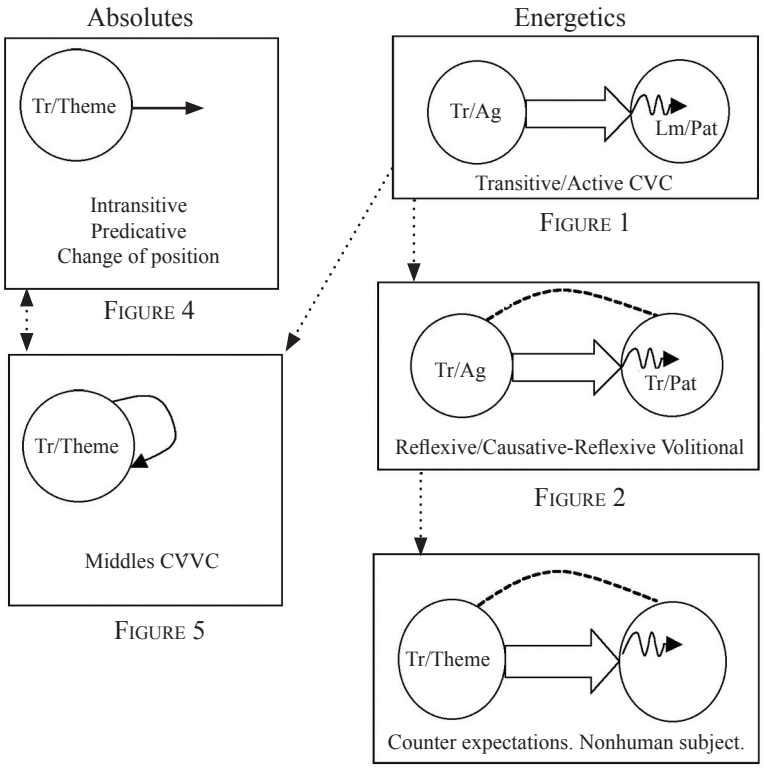


FIGURE 4

FIGURE 1

FIGURE 2

FIGURE 5

FIGURE 3

There has been a general tendency to assume that middle constructions develop historically from reflexives (Faltz 1985; Kemmer 1993, 1994; Lehmann 1995). As already shown in Nava & Maldonado (2005), this is by no means a universal. YM aligns with languages like Tarascan (Nava & Maldonado 2005), Toba (Mesineo 2002), Otomí (Palancar 2004), Balinese (Artawa 1994), and many others in which the middle did not develop from the reflexive construction. While for some languages the middle may be a basic (nonderived) voice, for others it may derive from an alternative base form. The case of YM is particularly interesting as it comes specifically from the transitive active construction. We have suggested in this paper that the middle offers the possibility of seeing the most energetic event in a neutral absolute manner.

While we were collecting data for this study, we were intrigued to find that for some very similar situations either an intransitive or a middle was acceptable, and that in others

either the middle or the reflexive construction could be used with no apparent restriction. Our fieldwork methodology was crucial for clarifying the system. Direct elicitation could only show us that certain events (e.g., healing) could be encoded as either middle or reflexive, but could shed little light on what guided a speaker's choice. From texts, we could find contexts suggesting the specific situation that would license one construction over the other. The same was true for change of position (e.g., kneeling and sitting would take a reflexive marker or an intransitive verb in very similar contexts). However, oral narratives suggested that the reflexive implied intentionality of the subject (e.g., hiding from a fox, looking at a pretty girl, etc.). On the other hand, no example of an emphatically volitional act could be found for the middle or the intransitive construction. Although our intuitions seemed correct, only the speakers' approval validated our claims. In follow-up elicitation, speakers not only approved our hypothesis, but they also created for us the contrastive contexts in which a reflexive could be used for the energetic version of an event, as opposed to the absolute, energy-neutral manifestation of non-reflexive constructions. We concluded that absolute and energetic construals in YM constitute two fundamental cognitive patterns that determine the behavior of intransitive constructions. As determined by the speaker's needs, the natural energy involved in an action may be highlighted to stress the subject's intention to attain a goal.

In recent years special attention has been paid to pragmatic phenomena codified in the grammatical organization of human languages. From two main perspectives put forward by Langacker (1987, 2000, n.d.), Traugott (1988, 1995, 2003), and Traugott & König (1991) it has been acknowledged that the view of the conceptualizer tends to determine the configuration of a grammatical construction. Whether the grammatical formation is determined by a shift from the referential world to the dominion of the speaker, or whether the conceptualizer enters the objective scene to impose a subjective view, what is crucial is that languages create mechanisms to describe situations as seen by the speaker. Middles in Yucatec Maya do not follow typologically common patterns of middle marking. Instead, they impose a view of events in neutral terms. This view is necessary in the system to contrast with the reflexive, a construction that gives special status to the subject's control and intentionality in driving the event. The importance of control for YM is corroborated by speakers' readings of reflexive constructions as accidental, obtained when the subject is unable to exert control over the action. Unlike many Mesoamerican languages that do not allow non-human subjects, YM exploits the reflexive construction to represent situations that contradict basic Maya cultural patterns. Middle, reflexive, and intransitive constructions alternate in YM to let speakers shape their view of an event. The data we were able to account for suggests that the interdependence between active, middle, intransitive, and reflexive constructions in YM not only makes a very efficient communicative system; it also shows that grammar and cognition are shaped in accordance with the richness of culture.

References

Andrade, Manuel J. & Hilaria Máas Collí. 1991. *Cuentos Mayas Yucatecos,* vol. I & II. Mérida: Universidad Autónoma de Yucatán.

Artawa, Ketut. 1994. Ergativity in Balinese Syntax. Melbourne, Australia: La Trobe University dissertation.

Ayres, Glenn & Barbara Pfeiler. 1997. *Los verbos mayas, la conjugación en el Maya Yucateco moderno.* Mérida, Yucatán: Ediciones de la Universidad Autónoma de Yucatán.

Bohnemeyer, Jürgen. 2004. Argument and event structure in Yukatek verb classes. In Ji-Young Kim & Adam Weerle (eds.), *The semantics of under-represented languages in the Americas* (SULA) (University of Massachusetts Occasional Papers in Linguistics 25). 8-19.

Briceño Chel, Fidencio. 2004. De voz en voz hay cambio de tono en Maya. In Zarina Estrada Fernández, Ana Fernández Garay & Albert Álvarez González (eds.), *Estudios en lenguas amerindias. Homenaje a Ken L. Hale*. Hermosillo: Universidad de Sonora.

Bricker, Victoria. 1981. Grammatical introduction. In Eleuterio Po'ot Yah (ed.) *Yucatec Maya verbs* (*Hocabá dialect*). xlix-xcii. Latin American Studies Curriculum Aids, Center for Latin American Studies. New Orleans: Tulane University.

Bricker, Victoria. 1998. *A dictionary of the Maya language as spoken in Hocabá, Yucatán.* Salt Lake City: The University of Utah Press.

Faltz, Leonard. 1985. *Reflexivization: A study in universal syntax*. New York: Garland.

Haiman, John. 1980. The iconicity of grammar. *Language* 56. 515-540.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59. 781-819.

Hill, Jane. 1969. Volitional and non-volitional verbs in Cupeño. *Chicago Linguistic Society* 5. 348-356.

Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam & Philadelphia: John Benjamins.

Kemmer, Suzanne. 1994. Middle voice, transitivity, and the elaboration of events. In Barbara Fox & Paul J. Hopper (eds.), *Voice: Form and function*, 179-230. Amsterdam & Philadelphia: John Benjamins.

Langacker, Ronald. 1987. *Foundations of cognitive grammar vol. 1: Theoretical prerequisites.* Stanford: Stanford University Press.

Langacker, Ronald. 2000. *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Langacker, Ronald. n.d. Subordination in cognitive grammar. Unpublished ms.

Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: LINCOM Europa.

Maldonado, Ricardo. 1988. Energetic reflexives in Spanish. *The Annual Meeting of the Berkeley Linguistics Society* 14. 153-165.

Maldonado, Ricardo. 1999. *A media voz: Problemas conceptuales del clítico se.* Location: Universidad Nacional Autónoma de México.

Maldonado, Ricardo. 2009. Middle as a basic voice system. In Lilian Guerrero, Sergio Ibáñez & Valeria Belloro (eds.), *Studies in role and reference grammar.* México: Instituto de Investigaciones Filológicas, UNAM.

Mesineo, Cristina. 2002. La marcación verbal activa/inactiva en toba (gauycurú) y sus motivaciones. (LIAMES (Revista de Lingüística Indígena americana) 2), 38-50. Campinas: Instituto da Linguagem, Departamento de Lingüística, UNICAMP.

Nava, Fernando & Ricardo Maldonado. 2005. Basic voice patterns in Tarascan. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture and mind,* 461-478. Palo Alto: Center for the Study of Language and Information & Stanford University Press.

Palancar, Enrique. 2004. Middle voice in Otomi. *International Journal of American Linguistics* 70(1). 52-85.

Traugott, Elizabeth Closs. 1988. Pragmatic strengthening and grammatical organization. *The Annual Meeting of the Berkeley Linguistics Society* 14. 406-416.

Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalization. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation*, 37-54. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs. 2003. From subjectification to intersubjectification. In Raymond Hickey (ed.), *Motives for language change*, 124-139. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs & Ekkehard König. 1991. *The semantics-pragmatics of grammaticalization revisited.* In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to gramaticalization, vol. 1*, 189-218. Amsterdam & Philadelphia: John Benjamins.

Israel Martínez Corripio
israim@hotmail.com

Ricardo Maldonado
msoto@servidor.unam.mx

8

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 173-202
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4455

Studying Dena'ina discourse markers: Evidence from elicitation and narrative

Olga Charlotte Lovick

Alaska Native Language Center, University of Alaska Fairbanks

This paper is concerned with discourse markers in Dena'ina Athabascan. One problem for transcribers and translators of Dena'ina texts is the great number of particles (i.e., words that cannot be inflected) that, according to speaker judgments "have no meaning" or "mean something else in every sentence." This suggests that these particles are discourse markers, whose function is to relate discourse units to each other and to the discourse as a whole. The paper contrasts two different forms of linguistic inquiry: direct inquiry in the field, by elicitation of meaning and function of the discourse markers, and indirect inquiry, by study of a corpus of Dena'ina narratives. While elicitation is helpful in obtaining an initial gloss for the discourse markers, it is shown that only the study of texts will give us insight into the function of such particles and allows us to understand the important differences between particles that, on first sight, appear to be synonymous.

1.INTRODUCTION. It is only recently that any aspect of Athabascan syntax above the clause level and any aspect relating to discourse have been tackled. At the time of this writing, several accounts exist for clause combining (Rice 1989, Saxon 1998, Kibrik 2004, Holton 2004, Fernald & Perkins 2006, Mithun 2008), and for discourse intonation (Holton 2005, Tuttle & Lovick 2007, among others). While most of these accounts are concerned with only one or two languages, the languages generally seem similar enough to be comparable. Thus there is now some basic understanding of these topics in Athabascan in general.

Several articles on cohesion and coherence (McCreedy 1989, Thompson 1989, Saxon 1993) focus on how the use of pronominal prefixes establishes topic continuity. The most inclusive discourse study is de Reuse & Mullin's (2005) study of discourse in Western Apache recipes; they discuss how pronominal prefixes, classificatory verbs, and discourse particles can combine to structure discourse. The present paper is a first attempt at a discussion of Dena'ina discourse markers.

Dena'ina (also known as Tanaina) is a Northern Athabascan language spoken in south-central Alaska surrounding the Cook Inlet and on both sides of the Alaska Range. The geographic diversity of the Dena'ina homeland is responsible for the great dialectal diversity; there are four major dialects with considerable variation even within the same major dialect (see Kari (1975) for a detailed discussion of Dena'ina dialectology). Dena'ina is severely endangered with only about 60 fluent speakers, most of them 60 years or older. Linguistic documentation of Dena'ina is available from the 1970s onwards, primarily through the efforts of James Kari and Joan Tenenbaum. The great dialectal diversity poses a severe problem for language revitalization measures, since materials have to be adapted to the dialect spoken in each community.

One problem for transcribers and translators of Dena'ina texts is the great number of particles (i.e., words that cannot be inflected) that, according to speakers I have worked with, "have no meaning" or "mean something else in every sentence." In the notebooks of Joan Tenenbaum (1973), for example, many particles (e.g. *idi ełа*; see section 4) have highly variable glosses, and several particles (*ha'* and *q'u*, discussed in section 3) are hardly ever glossed at all.

The assumption made here is that these particles are discourse markers in the sense of Schiffrin (1987:31ff.): "sequentially dependent elements which bracket units of talk," that is, lexical items whose presence cannot be explained just syntactically or semantically and that are instead best defined in terms of their discourse function, in particular, their ability to mark narrative paragraph breaks or the absence thereof. One subset of discourse markers, namely evidentiality markers, has received some attention (Rice 1989, de Reuse 2003, Holton & Müller 2005, Holton & Lovick 2008), but there are discourse markers with other functions that have not yet been discussed for any Athabascan language.

To define the meaning of one of these discourse markers, simple elicitation is not sufficient, despite its undisputable usefulness for initial glossing of lexical items. Direct questioning about the meaning of the particles considered here may or may not yield a useful answer. Take, for example, the conjunctions *ch'u, ch'q'u, ch'łuq'u*: a speaker will translate them with 'and'. Further elicitation will then show that these particles cannot be used to connect NPs, instead, the postposition *eł* is used following both nouns:

(1) (Lovick Fieldnotes, November 8th, 2005)

a. * *Mary ch'u/ch'q'u/ch'łuq'u John*

b. *Mary eł John eł*
Mary with John with
'Mary and John'

But only discourse analysis can show that *ch'u, ch'q'u,* and *ch'łuq'u* have different functions corresponding to their position within a narrative unit (line). Thus, the information gleaned from elicitation is supplemented by discourse information.

Similarly, the distribution of some of these markers in elicitation is fundamentally different from their distribution in narratives. For example, the marker *k'i* 'also' is particularly frequent in person paradigm elicitation. See (2):

(2) (Lovick Fieldnotes, November 9th, 2005)

a. *t-gh-esh-ch'ix*[1]
INCEP-FUT-1SG-breathe
'I will breathe'

[1] Abbreviations used here include: ADV='adverbial'; AREAL='areal prefix'; CLF='classifier'; CNJ='conjugation marker'; CONJ='conjunction'; DEM='demonstrative article'; DIST='distributive plural'; DM='discourse marker'; EVID='evidential'; FUT='future'; HA='high agentivity'; INCEP='inceptive';

b. *nen* *k'i* *du* *t-gh-i-ch'ix?*
2SG too QUEST INCEP-FUT-2SG-breathe
'will you also breathe?'

c. *yin* *k'i* *t-u-ch'ix*
3SG too INCEP-FUT-breathe
's/he too will breathe'

While I have hundreds of examples with *k'i* in my fieldnotes, it occurs in the corpus used for this study only 22 times (in 1,104 lines). It is thus not a particularly frequent discourse marker, which is not predicted from its frequency in elicited speech.

Drawing on a corpus of 15 narratives, I investigate the functions of several Dena'ina discourse markers. For this purpose, I will look both at elicited and at textual data. I will show how results from both kinds of linguistic analysis supplement each other, thus helping us to understand the meaning and function of Dena'ina discourse markers.

2. LANGUAGE, CORPUS, DEFINITIONS.

2.1. THE LANGUAGE. Athabascan languages are famous for their complex polysynthetic morphology, and Dena'ina is not an exception. For reasons of space, no morphemic glosses are provided beyond section 2 of this article. I realize that this may be inconvenient to readers interested in Dena'ina morphology, and refer those readers to Tenenbaum (1978) and Lovick (2006), both of which contain descriptions of Dena'ina morphology and numerous examples with morphemic glosses. In the context here, morphemic glosses add information that is irrelevant to the points made in this paper, and are thus omitted.

There is no comprehensive description of Dena'ina syntax available, so a brief discussion of this is necessary.

2.2. SYNTAX LIGHT. Dena'ina is a verb-final language (SV and OV, according to Dryer's (1997) typology), and only a few particles—typically evidentials, inferentials, and speech act markers—may follow the verb. There is the possibility of a NP or an adverbial following the verb, but these phrases are clearly marked prosodically as afterthoughts.[2] (3)-(5) are examples of typical Dena'ina word order.

(3) verb-final (*Quch' Nushjun Story*)
Kiq'u yethdi nitsinitsey.
kiq'u yethdi ni-tsi-n-i-tsey
again and.then up.to-head-CNJ-PF-move.elongated.object.quickly
'And another one stuck its head out.'

INDEF='indefinite'; NARR='narrative marker'; NEUT='neuter'; PF='perfective'; POSS='possessive'; QUEST='question marker'; REL='relative'; SG='singular'; TOP='topic'.

[2] See Tuttle & Lovick (2007) for details on Dena'ina intonation.

(4) with post-verbal evidential clitic (*Quch' Nushjun Story*)
Ch'adach' daghiłtey ghu shida.

ch'adach'	d-gh-i-ł-tey	ghu	shida
thus	NEUT-CNJ-PF-CLF-strong	there	EVID

'He was still that strong.'

(5) with afterthought indicated by comma (*Quch' Nushjun Story*)
Ch'u begguya k'iydlan, kił.

ch'u	b-gguya	k'-z-d-lan	kił
CONJ	HA-small	INDEF-CNJ-CLF-be	boy

'And he had a child, a boy.'

Tenenbaum (1978:29) gives the canonical word order preceding the verb as subject-object, and this is certainly true for elicited sentences. However, utterances with SOV structure are not very common in natural discourse, because very few utterances contain both an overt subject and an overt object NP. Not only is this typologically common (see Dryer (1997) for a discussion), but Dena'ina also has a large system of pronominal prefixes indexing subject and object arguments on the verb (see Tenenbaum (1978) for an inventory and Lovick (2006) for typical marking patterns). Thus, the use of free NPs is restricted to cases where pronominal reference is not sufficient, mainly the introduction of new referents (and related to that, change of referents).[3]

There are a few cases where both arguments of a transitive verb are co-indexed with free NPs. However, the order of the two NPs does not necessarily reflect their syntactic function but rather their relative animacy (Lovick 2006). Also, NPs marked by the new topic marker *-hdi* (see Section 3.1) are, independent of their syntactic function or their relative animacy, moved into the sentence-initial position. The topicalized NP can be co-referential with either the subject or the object:

(6) topicalized subject
Chida kuya gunhdi naghelt'ana łyes.

chida	kuya	gun-hdi	n-gh-l-t'a-na	ł-yes
old.lady	grandchild	DEM-TOP	DIST-CNJ-CLF-be-REL	CLF-save

'The old lady's granddaughter saved lots of people.'

(7) topicalized object
Shdaja ghunenhdi nen t'inluggen shughu.

sh-daja	ghunen-hdi	nen	t'-n-luq-en	shughu
1SG-younger.sister	DEM-TOP	2SG	thus-2SG-do-REL	EVID

'You [are the one] who did this to my sister.'

[3] In those cases, the pronominal prefix may or may not be present, see Lovick (2006).

The topic position in Dena'ina is at the beginning of the utterance. If the topicalized NP is co-referential with the subject, there is no change in word order, but if it is co-referential with the object, the NP is frequently (but not always) moved into the topic position, a process known as left dislocation.

2.3. THE CORPUS. For this study, a corpus of 15 traditional narratives comprising 1,104 lines was used. All narratives are oral and have been transcribed and translated by a team consisting of a linguist and a native speaker. For all narratives, I consulted the original audio, as discourse markers are sometimes omitted from transcriptions. No primarily written narratives (e.g. from the collections of Wassillie (1980) or Kalifornsky (1991)) were used, as they contain fewer-than-normal discourse markers. Traditional narratives were chosen because of the ready availability of the materials; many of them have been published, and some of them are available over the web at http://qenaga.org.

It had been planned to balance the narratives by dialect, but this proved to be impossible. The majority of Dena'ina materials are in the Inland and Upper Inlet dialects, and only a few sources exist for the Outer Inlet and Iliamna dialects. Thus there is no representative from the Iliamna dialect here, and only two short narratives in the Outer Inlet dialect. All narratives are listed in Appendix I.

Most of these narratives are traditional stories (the Dena'ina term is *sukdu*), with the exception of the Geese Story, the story of the Underwater People (the arrival of the first Russians in Dena'ina country), and the Story of Susitna, which are historical narratives. *Sukdu* are classic folk tales and leave a lot of freedom to the narrator—which explains why several versions even by the same narrator can differ as much as they do. They are not to be confused with more formal genres like poetry (see Tenenbaum (2006:xiv)), epos, or oratory. The focus on the narrative genre is mainly due to the availability of materials; much of the Dena'ina fieldwork by Kari, Tenenbaum, and myself has been centered on narrative and lexical data, and morphological description.

The narrative data is supplemented by my own fieldnotes (March 2005, November 2005, May 2006, and June 2006), which represent the 'elicited data'. The data collected on the first two trips (mainly verb forms and paradigms for educational purposes) first made it obvious that particles can occur in unexpected places in elicitations, and thus set the research agenda for the last two trips.

My observation during these interviews was (not surprisingly) that fieldwork on discourse markers is remarkably tricky. Asking direct questions "What does X mean?" was frequently unsuccessful and frustrating for the speaker. Several times I would present speakers with made-up sentences containing discourse markers, and they would latch onto all other parts of the sentence, ignoring the discourse marker; this of course would be frustrating for me. In the end, it proved most useful to discuss discourse markers on the side while overtly working on a task like translation or transcription, and to use material gathered in elicitation of verb forms etc. as a guide for asking further questions.

2.4. DEFINITION OF DISCOURSE UNITS. One problem for the study of discourse is the labeling of the units under discussion, and there are many complementary approaches. Chafe (1980, 1987) uses the terms 'intonation units' (defined by surrounding pauses), 'extended clauses' (intonation units organized into clauses or clause-like structures), and

'paragraphs' (cf. Chafe 1987:41f.). These three units are cognitive units, driven by "basic cognitive phenomena such as memory and consciousness." He lastly discusses sentences as belonging to the "category of phenomena which are under more rhetorical control, and more independent of cognitive constraints" (Chafe 1987:41).

Givón (1983), on the other hand, takes the clause as the basic information unit, assuming that one proposition is mapped onto one clause. Clauses then combine into paragraphs, and paragraphs into texts.

For the present analysis, the basic unit will be called a *line*. Lines are defined by speaker judgment during the transcription process. Lines correspond to Givón's clauses insofar as they are the basic information unit. In this sense, lines are semantic units. They are not necessarily syntactic units and may correspond to entities smaller or larger than, or isomorphic with, a syntactic clause. Prosodic cues that may correspond to the ends of lines are falling intonation, lengthening of the final syllables, and following pause (Tuttle & Lovick 2007). In this respect, they bear resemblance to Chafe's intonation units.

Narrative paragraphs (or episodes) can be defined and studied in two ways. One way has been suggested by Givón (1983:7): A paragraph is characterized by thematic continuity, action continuity, and topic/participant continuity, that is, a paragraph is taken to be an entity that is defined from within itself. This approach is used by McCreedy (1989), Thompson (1989), and Saxon (1993), by studying how topic continuity is expressed in units of text.

A paragraph can also be defined in relation to other paragraphs. Chafe (1987:42) notes several prosodic cues for paragraph breaks, such as "an increase in fumbling and disfluency, (...) where an interlocutor is especially likely to contribute some encouraging noise and remark." Assuming that thematic continuity, action continuity, and topic/participant continuity hold a paragraph together on the inside, it can be concluded (following Chafe (1980:40ff., 1987:42)) that a lack of one or all three of them signals a paragraph break. In Dena'ina, topic discontinuity and action (dis)continuity are frequently marked by discourse markers.

3. DISCONTINUITY.

3.1. NEW TOPIC MARKER *-HDI*. The marker *-hdi* occurs a total of 37 times in the corpus. Speakers tend to gloss it with 'how about' or 'as for', which suggests that it functions as a new topic marker. See (8):

(8) (*Two Women Story*)

1 *"Nen-**hdi**,"* *yełni* *łu*.
2SG-TOP she.said NARR
'"How about you?" [the one woman] said.'

2 *"Shi-**hdi*** *nuti'at* *hggagga* *tgheshlał,"* *yełni* *łu.*
1SG-TOP saltwater bear I'll.be she.said NARR
'"As for me, I will be a saltwater bear," [the other woman] said.'

While topic continuity is one of the main criteria of paragraph cohesion (see Givón (1983) for a typological study, McCreedy (1989) on Navajo, and Thompson (1989) on Koyukon and Navajo), topic discontinuity is a good criterion for a paragraph boundary.

Following Givón (1983:8), I define *topic* as the "participant most crucially involved in the action sequence running through the paragraph."[4] Once a topic has been established, topic continuity is usually indicated by pronominal reference to a nominal antecedent (cf. Thompson (1989) for a discussion of topic continuity in Navajo and Koyukon, among other languages). In one rather spectacular Dena'ina example, the main referent is not referred to by an overt NP for 36 lines. Topic discontinuity, however, is frequently marked by the enclitic *-hdi*, a 'new topic' marker. *-hdi* always attaches to the last element of the NP, not necessarily to the head noun; in the example below, it attaches to the demonstrative article *ghun*, rather than to the head noun *qeshqa* 'rich man'.

The following example is from the Crane story. The rich man and his daughter treat the cranes badly and are punished for this. In (9), the cranes are flying overhead, announcing bad weather (the punishment). The rich man does not read the signs correctly. He assumes that the weather is going to be good and that that is happening because of him. As a consequence, nearly everyone dies from the cold.

(9) (*Crane Story*)

1	*Hch'anindatl'*	*ch'u*	*dilghuy*	*kegh*	*gheli.*			
	they.flew.out	CONJ	they.called	big	really			
	'[The cranes] flew out and called loudly.'							
2	*Qek'di'un.*							
	the.sun.came.out							
	'The sun came out.'							
3	*Qeshqa ghun**-hdi***	*qughiyu*	*ch'u*	*tał*	*uch'en*	*t'ugh*	*nidalkits.*	
	rich.man DEM-TOP	he.came.out	CONJ	mat	outside	down	he.laid	
	'The rich man came out and laid a mat outside on the ground.'							
	[two lines of direct speech omitted]							
4	*"Shi*	*shghuda*	*shughu*	*t'ant'i,"*	*yełni.*			
	I	because.of.me	EVID	it.is	he.said			
	'"This good stuff is happening on account of me," he said.'							

[4] I am not concerned with grammatically coded sentence topics as described by Givón.

5 *Tałq'e* *qedełtan.*
on.the.mat he.was.lying.down
'He was lying on the mat.'

6 *Ndał* *yeh* *hch'anindatl'* *ghuda* *shughu* *nch'uk'a* *qit'aniyen.*
cranes there they.flew.out because EVID not he.knew
'The cranes flew over, but he didn't know the real reason.'

7 *Uch'en* *dałch'ey* *kegh* *gheli.*
outside windy big really
'Outside the wind started to blow really hard.'

8 *K'itigi* *edli* *ch'u* *qeshqa* *yugh dunu'iltlet.*
too.much cold CONJ rich.man he.ran.back.inside
'It was really cold and the rich man ran back inside.'

The first two lines of (9) are the last lines of a paragraph dealing with preparations against the coming cold weather. In line 3, the rich man, topic of the next paragraph, is re-introduced into the narrative, marked by *-hdi*.

This function of *-hdi* is probably responsible for its pervasiveness in person paradigm elicitations: The frequent switches are an example of highly unnatural topic discontinuity, and some speakers feel the need to mark this.

(10) is taken from the beginning of the Grayling story. Lines 1-5 are the introduction; they give the background to the story of how a woman refuses to pick berries and turns into a grayling for it. The woman is introduced into the narrative in line 6. The topic marker *-hdi* signals not just that a there is a new topic, it also signals the beginning of the main action in a new paragraph.

(10) (*Grayling Story*)

1 *Dehghelch'ehi* *ey* *hehdułał,* *degget qenen* *qahnidatl'.*
baskets there they.carried uphill they.went
'They were carrying their baskets and they went up on the hillside.'

2 *Nek'qiniya.*
they.were.picking.berries.in.places
'They were picking berries in places.'

3 *Nek'qiniya,* *yudeq* *huqudeł.*
they.were.picking up they.were.going
'They were picking and they were going up.'

4 *Yudeq* *huqezdatl'* *hna,*
up they.went while
'While they went up above,'

5 *k'qiniya,* *t'anch'q'u* *k'qiniya.*
they.were.picking everyone they.were.picking
'they were picking, everyone was picking.'

6 *Ts'ił t'an* *deghk'isen* *gunen-**hdi*** *łu*
one woman DEM-TOP NARR

qeykuh *hk'uch'* *zdu.*
below.them down.there she.stayed
'This one woman is staying away and was down below away from them.'

7 *"K'qiniya"* *qeyłnih* *hq'u* *nch'u* *k'inesya.*
you.pick they.tell.her but not she.picks.not
'"You pick," they say to her but she was not picking.'

Assuming that paragraphs are, among other things, characterized by topic continuity, it follows that a clitic marking topic discontinuity can be used to signal the beginning of a new paragraph. In nearly all of the 36 occurrences in the corpus, *-hdi* cliticizes to an NP (re)introducing a new topic. Rarely, it can attach to a VP:

(11) (*Caribou Story*)

1 *Ch'u* *q'uyehdi* *nigiga* *yet* *yinunetneł* *iqech'* *ghu*
CONJ and berries there she.poured.them.in.it thus there

t'eyeł'an *t'eyeł'an* *ch'u.*
she.did.this she.did.this CONJ
'Then she poured the berries into it big bag and she did this and did this.'

2 *Q'uyehdi* *yidak'itnaniya.*
and.then she.picked.so.it.was.full
'Then she had picked until this was full.'

3 *Łu* *q'uyehdi* *ey* *q'ut'un teh* *hch'anadyux* *ch'q'u*
NARR and.then there every.morning she.goes.back.out CONJ

łu *vejex* *uqu hnił'an,* *k'qisen.*
NARR caribou she.looks.for there.is.nothing
'Every morning she goes back out to that flat and she looks for caribou, but there's nothing.'

4 *Kiq'u* *nuk'inedya**-hdi*** *yinuk'eneł* *yi* *giga* *yinunetneł.*
again she.picked.berries-TOP she.poured.in there berries she.poured.in
'She picked berries again and she poured them in again, she poured in more berries.'

5 *Kiq'u* *yet* *hunusdyu* *idi ełа* *łu* *vejex.*
again there she.walked.up CONJ NARR caribou
'Again she walked up there and then caribou.'

The stretch of text in (10) describes how a woman prepares to deal with the (fanged) caribou; she fills a large sack with berries and climbs in so that the caribou will bite their teeth out on the sack without hurting her. Lines 1-3 describe the preparations, and how she would look out for caribou. Line 4, where she is again picking berries, leads over to line 5, where she finally sees caribou.

Following a VP, *-hdi* does not have paragraph breaking function and instead seems to indicate background action. However, there are too few (three) examples in the corpus to be certain.

3.2. NEW ACTION MARKER *Q'U*. *Q'u* is one of the most frequent particles in the corpus with 116 occurrences. *Q'u* is so pervasive in Dena'ina discourse that it is very hard to assign a meaning to it—it seems to have so many different functions that it is hard to see how they all relate to each other. In fact, it frequently is not glossed at all because its meaning is so hard to determine.

In the Dena'ina stem list, Kari (n.d.) glosses it with 'now' and 'emphatic', and points out that it commonly occurs in adverbials which specify a point in time or emphasis. Speakers gloss it with 'now' in the rare cases where they gloss it at all—frequently, they claim that it "has no meaning."

In elicited sentences, *q'u* is remarkably rare and almost never has the meaning 'now'. If the meaning 'now' is intended, speakers will usually prefer the adverb *q'udigu* 'right now':

(12) (Lovick Fieldnotes, November 8th 2005)
Q'udigu *qayeh* *qech'* *ghesyuł.*
right.now village to.it I'm.walking
'I'm walking to the village right now.'

More common in elicitation is the 'emphasis' reading of *q'u*. It occurs particularly frequently in questions, following and emphasizing the question word:

(13) (Lovick Fieldnotes, May 19th 2006)
Nunkdahdi *daha* ***q'u*** *eydu?*
your.mother.TOP where EMPH she.lives
'Where does your mother live?'

Both functions of *q'u* also occur in connected speech. In this paper, I will concentrate on the 'now' reading of *q'u*, because of its clear discourse functions. Brief mention of the 'emphasis' reading of *q'u* will be made at the end of this section.

The gloss 'now' suggests that *q'u* is a temporal adverb referring to either the reference time (the real-world time when the line is uttered) or to the event time (a point in time within the narrative; the terminology follows Schiffrin (1987)). The first usage, shown in (14), is straightforward and can be observed several times in the corpus:

(14) (*Quch' Nushjun Story*)
Q'u *dutgheshcheł.*
ADV I.will.shut.up
'I will shut up now.'

The point of time the speaker refers to is set in the real world and not the story world, the story-telling situation rather than an event in the story itself. The example in (14) is taken from the end of the story, where the speaker transitions from the storytelling to the real world. Due to the nature of the corpus (narratives as opposed to conversation), cases like this are fairly rare.

(15) (*Quch' Nushjun Story*)

1	*En'ushen*	*gheli*	*iydlan*	*tets'*	*ghini*	*belaq'a*	*ich'a*
	old.man	really	he.became	spear	DEM	his.hand	from.it

dinghel.
it.dropped
'He got real old and that spear dropped out of his hand.'

2	*Ch'u*	*guhqugh*	***q'u***	*dagheshłggech'*	*t'eghetl'ił.*
	CONJ	this.far	ADV	I.make.it.short	I'm.going.to.do.it

'Here I'm going to shorten the story.'

3	*Ch'u*	*qubeł*	*nuhghiq'uts'.*
	CONJ	to.them	it.got.cold.again

'And it got to be fall time on them.'

Lines 1 and 3 of (15) contain the main story line, and line 2 is a comment of the speaker on the telling of the story, rather than on the story itself. Again, the storyteller transitions from the story world to the real world—only this time, he transitions back in the following line.

Somewhat trickier is the second usage, *q'u* referring to a point of time within the event time. A straightforward example is (16):

(16) *q'u* referring to event time (*Hunting Dog Story*)

1 *Łik'a ghin, hełch' qubegh qiydlan*
dog DEM evening to.them it.became

ki q'u qeyech' tunudalggey.
again they.walked.away.from.him
'When it became evening, they again walked away from the dog.'

2 *Bingha ghun ki biydełchih gheli.*
his.older.brother DEM again he.was.scolding.him really
'His older brother was really bawling him out.'

3 *"Łik'akda ghin yach' tiłt'uh* ***q'u,"*** *yełni.*
old.dog DEM away you.throw.him ADV he.said.to.him
'"Throw away that old dog!" he said to him.'

4 *"**Q'u** hełch' na'eł taydlan."*
ADV evening to.us it.has.become
'It has become evening on us now.'

5 *Jani hqugh enhghiydlu.*
all.day.long they.had.been.staggering.along
'All day long they had been staggering along.'

Both occurrences of *q'u* in (16) refer to points of time within the narrated time. The text is about two brothers who are trying to find food for their family. A dog is following them. The younger brother feeds and carries the dog, while the older brother keeps encouraging him to get rid of it. The paragraph above describes the second time this occurs (hence *kiq'u* and *ki* 'again' in lines 1 and 2). The older brother is getting very annoyed and insistent (line 3), and accuses the younger brother of wasting valuable time (line 4).

Frequently, *q'u* does not just indicate that a certain event is taking place at a particular time within the story. Instead it functions as a discourse marker (rather than as temporal adverb), signaling that a new action is about to begin. In this way, *q'u* marks the beginning of a new paragraph, and it is quite similar to English *now* as described by Schiffrin (1987:232). See (17).

(17) (Beginning of *Ch'iduchuq'a Story*)

1 *Ch'iduchuq'a gun lu.*
Ch'. DEM NARR
'This is Ch'iduchuq'a.'

2 *Tuzdatl'na* *guna* *łu.*
they.going.up DEM NARR
'Some people went up into the mountains.'

3 *Qunsha* *iqu* *qel'an* *dghiliq'* *hdalts'i*
ground.squirrels for they.trapped on.the.mountain they.stayed

ch'łuq'u *k'iqu* *qel'an* *k'iqu* *qel'an.*
CONJ again they.trapped again they.trapped
'They were trapping squirrels, they were staying up on the mountain and they kept trapping and trapping.'

4 *Oh,* *shan* *daghisedi* *ghin* *łu*
oh summer it.lasting DEM NARR

qunsha *iqu* *qel'an* *hnuyu.*
ground.squirrel for they.trapped when
'Oh, all summer long they trapped ground squirrels.'

5 ***Q'u*** *idi ełа* *eya'* *qunsha* *ghini* *nał* *qisil*
ADV CONJ oh.dear! ground.squirrel DEM on.us they.disappeared

qunsha *nch'u* *chihdełt'ik'* *ha't'qidyuq.*
ground.squirrel not they.killed thus.it.happened
'All of a sudden, oh! The ground squirrels disappeared on us, they didn't kill any more ground squirrels.'

Lines 1-4 of (17) offer background information; they provide a temporal setting (summertime) and they classify the story as a ground squirrel story: a story told in the summer, when people go up the mountains to trap squirrels, and at the same time a story that has squirrel trapping as its topic. The beginning of the main action of the story in line 5 is signaled by *q'u*. In this instance, *q'u* does not refer to any specific time; it instead situates the following events at an arbitrary point of time, the time that the story took place. All events narrated in the story occur in the time span following the point set by *q'u*.

Somewhat different is the use of *q'u* in (18).

(18) (*Crane Story*)

1 *Chida* *kuya* *nqela* *nqela* *ch'u*
old.lady granddaughter she.dug she.dug CONJ

qeshqa *uts'enaqa* *yet* *qeyaninudeł.*
rich.man his.children there they.came
'The old lady's granddaughter kept digging and digging and the rich man's children came over there.'

2 *K'eldunna* *qeynunetuh.*
some they.helped.her
'Some of them helped her.'

3 ***Q'u*** *q'aqinilu* *yethdi* *ndał* *ghin* *yeh*
DM she.finished.digging and.then cranes DEM there

hts'anindatl'.
they.flew.overhead
'Just as she finished digging, the cranes flew over.'

The first two lines in (18) describe how the girl digs a hole to protect her from the cold that the cranes are going to bring. They provide the background against which line 3 is set: again, *q'u* situates the following events. The difference in (17) is that the point in time set by *q'u* is not arbitrary but instead part of the story: the end of the preceding event (digging) is simultaneous with the beginning of the next one (cranes flying).

This is an important function of *q'u*: by situating events within the narrative, it can be used to define narrative paragraphs.[5]

(19) (*Quch' Nushjun Story*)

1 *Kiq'u ki* *yet* *nitsik'enitsey* *ka'a* *kda*.
again there something.stuck.its.head.out big.one indeed
'Another one put its head out, a big one.'

2 *Bejil'u* *qughettl'et* *hnuq'u* *he know brown bear*
its.ears they.came.out when

chił'ishen *quht'ana* *ghila* *ch'u.*
he.killing.them man it.was CONJ
'When its ears came out, it knew that this was a man that kills brown bears.'

3 *Ggagga* *chich'eł'ishi* *quht'ana* *ghila.*
brown.bear he.killing.them man it.was
'He was a bear killer.'

4 ***Q'u*** *yeł* *tsadałnen.*
DM with.it he.clubbed.him
'He clubbed that one [with his tomahawk].'

[5] An anonymous reviewer pointed out that it is not clear whether *q'u* in (18) ends the preceding or begins the next paragraph. In cases like (18), this is hard to determine, possibly indicating that a paragraph break does not have to be totally rigid, and that in fact a line itself may constitute the break.

(19) is an interesting example because it contains a parenthetical. Lines 1 and 4 are the main story line (Quch' Nushjun encounters a bear and tries to kill it), but lines 2 and 3 contain the thoughts of the bear: the bear realizes that Quch' Nushjun is a bear killer. The initial *q'u* in line 4 signals the end of the parenthetical and indicates that the main storyline is resumed.

In (20) below, Chickadee has just told the women that their husband, whom they believed to be dead, is in fact living with another wife in a different village. What follows is their reaction to this discovery. Again, *q'u* signals the beginning of a new paragraph.

(20) (*Two Women Story*)

1 *"Hhį!"* *qeyłni* *łu.*
is.that.so! they.told.him NARR
'"Is that so!" they said to him.'

2 *"Yagheli,"* *qeyłni* *łu.*
good they.said.to.him NARR
'"Alright," they said to him.'

3 *Ch'ggagga* *gin* *nudnilen.*
chickadee DEM he.flew.away.again
'Chickadee flew away again.'

4 *Q'ut'un* *q'ut'un* *gheli* ***q'u*** *t'sahnizet.*
morning morning really DM they.woke.up
'Early in the morning, they woke up.'

5 *Nlugha* *huk'es'uk* *hq'u*
not.yet the.sun.hadn't.risen but
'The sun hadn't risen yet, but'

6 *ts'ahnizet* *ha'.*
they.woke.up CONJ
'they woke up.'

7 *Nuhuqestl'in* *nutihna* *qilan* *ch'q'u.*
they.got.dressed both they.were CONJ
'They got dressed, both of them.'

In (21) and (22), *q'u* 'now' also signals the beginning of a new paragraph. In both cases, *q'u* appears in the middle of the utterance, following an utterance that provides background information.

(21) (*Porcupine and Brown Bear*)
Tayanq' *hqugh* *sht'a* *yeł* *nitsighalkit*
middle as.far.as just with.him he.was.swimming

idi eł ***q'u*** *nini* *ghini* *yiłt'eh.*
CONJ DM porcupine DEM it.clubbed.him
'He had just swum out to the middle when the porcupine clubbed him.'

(22) (*Porcupine and Brown Bear*)
Yudeq *sht'a* *ni'ilgguk* *idi eł* ***q'u*** *ggagga* *yeh* *ch'aniyu.*
up just he.had.gone CONJ DM brown.bear there it.came.out
'He was just up when a brown bear came out.'

Q'u can also be used to add emphasis to question words (see (13) above) and noun phrases. See (23):

(23) (*Quch' Nushjun Story*)
1 *Yeqech'* *beł quyeltlet* *k'a* *nch'u* *t'inil.*
thus it.tried.to.jump.out.to.him too not it.couldn't.make.it
'It [bear] had tried to leap out to him [Quch' Nushjun] but couldn't do it.'

2 *Q'uyethdi* *yet* *nitsinitsey.*
and.then there it.stuck.its.head.out
'So it thrust its head there.'

3 *Chin shla* ***q'u*** *ulaq'a* *nuyu'uł* *ulach' gheli hqugh q'u*[6]
axe little **EMPH** his.hand he.grabbed.it.again with.all.his.might

yetsighala *kidghiłnen.*
top.of.its.head he.hit.it
'He grabbed his little tomahawk again and hit that bear on the top of the head as hard as he could.'

(23) describes a bear hunt. In the utterances preceding this stretch of text, Quch' Nushjun, who is looking for wood for sled crossbraces (hence the axe), comes across a bear den and starts killing the bears. At this point of the story, Quch' Nushjun is already an old man and not as strong as he used to be. However, he is still strong enough to kill a grizzly bear with one stroke, armed with just an axe, rather than with his special bear-hunting spear. The axe is emphasized this once and then not mentioned again throughout the text. *Q'u* thus emphasizes a referent once, but this emphasis is restricted to the line it occurs in.

[6] In this case, *q'u* is part of the complex adverbial *ulach' gheli hqugh q'u* 'with all his might, completely'.

To summarize, *q'u* has several functions. It can function as a local emphasis marker. It can also function as temporal adverb meaning 'now', but its function can be more than just situating an event in time. It also marks action discontinuity. There are three contexts where this happens particularly frequently: the background action is finished and the main action starts, one (old) action is finished and a new action starts, or a narrative digression is finished and the main action starts again. In all cases, *q'u* signals a coming action break, just as *-hdi* signals a coming topic break.

4.CONTINUITY.

4.1. CONJUNCTIONS: POSITION AND FUNCTION. Dena'ina has a remarkable number of conjunctions that can be roughly translated as 'and': *eł, ch'u, ch'q'u, ch'łuq'u, ha/ha',* and *idi eła.* Of these, *eł* conjoins only NPs.[7] Following Holton's (2004) analysis of Tanacross conjunctions, I originally assumed that different conjunctions have different scope, so that one conjunction would ideally connect clauses, a different conjunction would connect lines, and a third would maybe connect paragraphs. However, this assumption was not borne out by the data, and speakers were always ready to agree to substitutions of one conjunction with another (as long as they were dialectally appropriate, see section 4.2). Instead, it turned out that the *position* of a conjunction within a line indicates which units are connected. The co-existence of several synonymous conjunctions seems to be due to dialectal differences and very slight semantic differences.

Clauses within an utterance are usually conjoined with *ch'u, ch'q'u, ch'łuq'u* (henceforth *ch'u* refers to all variants) or *ha/ha'*.

(24) (*Mouse Story*)

Q'uyehdi	*yegh*	*n'ilgguk*	***ha'***	*yenił'an.*
and.then	to.her	he.went	CONJ	he.looked.at.her

'And then he went over to her and looked at her.'

(25) (*Ground Squirrel Story*)

Tinuhqenlyit	***ch'q'u***	*ndunuhqenlyit*	***ch'q'u***
they.keep.running.back.out	CONJ	they.keep.running.back.in	CONJ

yuyeh	*nch'u*	*hdilts'ik.*
inside	not	they.don't.stay

'They keep running back out and they keep running back in, and they don't stay inside.'

Lines can be connected using *ch'u* or *idi eła* in line-final position. In that position, *ch'u* and *idi eła* form an intonational unit with the preceding word (usually a verb).

---

[7] In other Athabascan languages, the postposition *eł* and its cognates (*ʔił* in Upper Kuskokwim Athabascan (Kibrik 2004), *eł* in Tanacross (Holton 2004)) can be used to coordinate clauses. This use of the postposition *eł* does not occur in the corpus. Also, Dena'ina speakers refused sentences with *eł* as a clause connective.

The use of *ch'u* and *idi ełа* in line-final position is a rhetorical means to keep the hearer interested. (26) below is taken from the *Two Women* story. Two women have been deserted by their husband. They turn into bears and get ready to attack the village where the husband now resides. In (26), the husband prepares his bow to shoot the two bears before he realizes that they are in fact his wives.

(26) (*Two Women Story*)

1 *Ts'iłten* *k'niq'nughuyel* ***ch'q'u***
bow he.nocked.an.arrow CONJ
'He nocked an arrow in his bow.'

2 *ts'ełq'i ghin* *nunch'* *nutastdyu*
one broadside she.started.walking
'one of [the bears] walked by him exposing her flank and'

3 *idi ełа* *ts'iłten* *ełа* *deqtalghel* ***ch'q'u***
CONJ bow with he.raised CONJ
'Suddenly he raised his bow and'

4 *ts'iłten* *ya nichet*.
bow he.pulled.it.back
'he pulled back the bowstring.'

The pace of the story in (26) is slowed down considerably, creating suspense ("when will the husband realize that he is about to shoot his wives?"). The intonation in each line in (26) indicates the end of a line, but the use of final *ch'q'u* indicates that more is to follow. (27) illustrates that the same is true for *idi ełа.*

(27) (*Caribou Story*)

1 *Ch'u* *yegh* *q'u* *q'u* *nughiłghatl'* ***idi eł***.
and to.her ADV ADV it got dark CONJ
'And then it got dark on her, and

2 *yech'ana'il'uch'*.
they migrated away from her
'the caribou herd moved away from her.'

Thus, *ch'u* and *idi ełа* create cohesion within stretches of text.

In sentence-initial position, *ch'u* and *idi ełа* have yet again a different function (*ha/ ha'* cannot occur there). They are used to connect paragraphs, rather than connecting lines within paragraphs. (28) is from the same source as (26), about 20 lines further into the story:

(28) (*Two Women Story*)

1 *K'elduna qut'ana chihdghiłt'ik k'elduna shtuqehnaniłyit.*
some people they.killed some they.ran.away
'They killed some people and some ran away.'

2 *Qayeh qayeh qizdlu t'anch'q'u qanqehnazchet ha'.*
house house AREAL.are all they.destroyed.them CONJ
'They destroyed all the houses in the village.'

3 *T'anch'gheli qanqehnazchet.*
all.of.them they.destroyed
'They demolished every last one of the houses.'

4 ***Ch'q'u*** *hch'a hch'a naqidyu.*
CONJ they.started.back.uphill
'Then they started back uphill.'

5 *Yudeq ghu k'ehnuldatl' ghu hunuqesdyu.*
uphill there they.ate.berries there they.went.back.there
'They got back up to the top, where they had eaten berries.'

Lines 1-3 are within the same paragraph. The beginning of the next paragraph (lines 4 and 5) is signaled by *ch'q'u*.

(29) shows that *idi eła* can also be used to start a new paragraph:

(29) (*Mouse Story*)

1 *Dlin'a shla gin eyeh nushełdułtex.*
mouse little DEM inside he.was.running.around
'The little mouse was running around inside.'

2 *Nalqeni łach yitse ndghiłtl'it'.*
hot ashes over.him he.poured
'He [man] poured hot ashes over him [mouse].'

3 *Really nalqeni.*
hot
'Really hot.'

4 *Yeh shtunushełdaltuk'.*
there he.ran.away
'He ran away from him.'

5 *Hey* *gheli* ***idi eła*** *k'undet* *tqedyuq.*
winter really CONJ starvation it.happened
'Really late that winter, they were starving.'

In this example, lines 1-4 belong to one paragraph and line 5 to another one. Lines 1-4 are set in the summer in the smokehouse (at the time when people prepare and store—or *put up*—fish for winter). They describe how the rich man's son hurts the mouse by pouring hot ashes over him. In line 5, the punishment for this deed is described: starvation sets in. The beginning of a new paragraph is signaled by the adverb *hey gheli* 'in late winter', but also by the connective *idi eła*.

Thus, *ch'u* and *idi eła* have three functions, depending on their position: They conjoin clauses in line-medial position; they conjoin lines in line-final position, creating tension, and they conjoin paragraphs in line-initial position.

4.2. CONJUNCTIONS: DIALECTOLOGY AND MEANING NUANCES. The question arises why there are several connectives with such similar meaning and function. One reason is dialectology. Table 1 shows which connective occurs in which dialects. Note that Lime Village and Nondalton are both Inland dialect. Speaker MH shows characteristics of both these villages (she was born in Lime Village, but has lived in Nondalton for more than 70 years).

TABLE 1. Connectives in the corpus (numbers refer to how frequently each connective occurs in the corpus)

	Nondalton	MH	Lime	Kenai	Upper Inlet	total
ch'łuq'u	5					5
ch'q'u	42		6		3	51
ch'u		5	20	14	108	147
ha'	18	11	3			32
idi eła	38	4	8	1	3	54

Ch'u is the preferred connective in Lime Village, the Kenai and in the Upper Inlet dialect. Speakers from Lime Village use both *ch'u* and *ch'q'u*.

Ha/ha', ch'łuq'u, and *ch'q'u* are used mainly in Nondalton. Of these, *ch'łuq'u* usually connects clauses, *ha/ha'* usually connect clauses and lines, and *ch'q'u* can connect clauses, lines or paragraphs. Kari (p.c.) points out that *ha/ha'* may be on the way to replacing *ch'q'u*, and elicitation work done in recent years has confirmed this; most present-day speakers prefer *ha/ha'* as clause and line connector (but will also accept *ch'q'u*). In a more vital language situation, this split between connectives might continue, resulting in a clear system like the one described by Holton (2004) for Tanacross, with different connectives indicating different degrees of closeness.

Idi ełа is the only connective that occurs in all dialects, although it is much more frequent in the Inland dialect. *Idi ełа* differs from the other connectives in its meaning; it is not a plain 'and' connective, but it can contain an element of surprise and suddenness. Indeed, speakers frequently gloss it with 'suddenly', even in cases like (29) above, where the onset of starvation in late winter is probably not sudden, but rather a gradual process.

(30) (*Mouse Story*)

1 *Qeshqa vey'a gun chaqenq'a yeh nugheyuł.*
rich.man his.son DEM smokehouse there he.was.walking.around
'The rich man's son was walking around in the smokehouse.'

2 *Kalajeł gini fish shla vendenghalts'etl'*
fish.drying.pole DEM little it.had.stuck.onto.it

nuydełdeł ch'u.
he.nibbled.on.that CONJ
'He chewed little pieces of fish that had dried on the fish pole.'

3 *He, he chew on, on the kalajeł.*

4 ***Idi'ełа*** *gu veghe ndu'ilggugen qilan.*
CONJ then to.him someone.coming.in there.was
'Suddenly, someone came in to him.'

In the story, (30) immediately follows (29) above; the smokehouse is the place where the rich man's son has poured hot ashes on the Mouse. Now he returns to this place because he is hungry enough to chew little bits of fish from the fish drying poles. The person entering in line 4 is of course the Mouse, who will explain to him about appropriate behavior towards animals. *Idi'ełа* (a variant spelling of *idi ełа*) indicates that the Mouse's appearance is sudden and surprising.

In this section, I have discussed five connectives, all of them with the meaning 'and'. I have been concerned with their exact function (which size of unit they connect) and meaning. I have seen that *ch'łuq'u* can only connect clauses. *Ha/ha'* can connect clauses and lines and may be gradually replacing *ch'q'u* with that function in the Inland dialect. *Ch'u* and *ch'q'u* can connect clauses, lines, and paragraphs, depending on their position within the line. *Ch'q'u* is used mainly in Nondalton, *ch'u* in all other dialects. *Idi ełа* can connect lines and paragraphs, depending on its position within the line. It occurs most frequently in line-initial position and is then usually glossed as 'suddenly'.

4.3. SEQUENTIALITY OF EVENTS: Q'UYEHDI AND YETHDI. There are two adverbial connectives in Dena'ina that can be glossed with 'then': *yethdi* and *q'uyehdi*. The main difference between them is dialectal with *q'uyehdi* being used mainly (but not exclusively) in the Inland dialect, and *yethdi* in the Upper Inlet and Kenai dialects. There is apparently no meaning difference. Thus, the discussion will treat both connectives together.

Q'uyehdi/yethdi are typically found at the beginning of a line. They usually occur when the events described are clearly sequential.

(31) (*Quch' Nushjun Story*)

1 *Yeqech' beł quyeltlet k'a nch'u t'inil.*
thus to.him it.tried.to.jump.out too not it.could.not
'It [bear] had tried to leap out but couldn't do it.'

2 *Q'u* ***yethdi*** *yet nitsinitsey*.
ADV then there it.stuck.its.head.out
'So it thrust its head there.'

[3 lines omitted, describing the killing of the bear]

3 *Kiq'u ki yet nitsik'enitsey ka'a kda*.
again again there something.stuck.its.head.out big indeed
'Another [bear] put its head out, a big one.'

[3 lines omitted, describing the killing of this bear]

4 *Yet yelugh yeł qiteł ch'u dutsi yeqala dedulnen*.
there its.end with.it it.jumped and on.top its.dirt.pile it.fell
'That bear jumped over it and it fell down over its pile of dirt.'

5 *Kiq'u* ***yethdi*** *nitsinitsey*.
again then it.stuck.its.head.out
'Then another one stuck its head out.'

At this point in the story, Quch' Nushjun has found a bear den with three bears in it and he starts killing them, one after the other. This is the kind of situation where *q'uyehdi/yethdi* tend to occur: when events sequentially follow one another, rather than when events overlap or occur simultaneously.

Q'uyehdi/yethdi do not have a clear paragraph breaking function, instead, they generally indicate action continuity. This is shown in the next example:

(32) (*Caribou Story*)

1 *Ch'u* ***q'uyehdi*** *nigiga yet yinunetneł iqech' ghu*
and then berries there she.poured.them.in thus there

t'eyeł'an t'eyeł'an ch'u.
she.did.it she.did.it and
'Then she poured the berries into a big bag and she did this and did this.'

2 ***Q'uyehdi*** *yidak'itnaniya*.
then she.picked.up.to.its.mouth
'Then she had picked until this was full.'

3 *Łu* ***q'uyehdi*** *ey q'ut'un teh* *hch'anadyux* *ch'q'u*
NARR then every.morning she.goes.back.out and

łu *vejex* *uqu hnił'an,* *k'qisen.*
NARR caribou she.looks.for there.is.nothing
'Every morning she goes back out to that flat and she looks for caribou, but there's nothing.'

4 *Kiq'u* *nuk'inedyahdi* *yinuk'eneł* *yi* *giga*
again she.picked.TOP she.poured.them.in in.it berries

yinunetneł.
she.poured.them.in
'She picked berries again and she poured them in again, she poured in more berries.'

The stretch of text above clearly is part of the same paragraph; the woman in the narrative is preparing for when she will meet the caribou, and all of the story units in this paragraph describe parts of her preparations.

Thus, *q'uyehdi/yethdi* 'then' has a markedly different function from adverbs like *q'u*, connectives like *ch'u*, and the topic marker *-hdi*. In contrast to all of these, *q'uyehdi/yethdi* indicate that there is *not* going to be a paragraph break of any sort, that the story unit following *q'uyehdi/yethdi* is part of the same paragraph as the preceding one. In this way, *q'u* and *q'uyehdi/yethdi* complement each other. *Q'u* starts a new paragraph, while *q'uyehdi/yethdi* continue a paragraph.

5. CONCLUSION. In the preceding sections, several discourse markers—a topic marker, several conjunctions, and two adverbs—have been discussed, and their discourse functions have been demonstrated. For the conclusion, I examine a longer stretch of text in order to show how the above-mentioned markers interact.

The beginning of the Raven story "Raven and his Two Wives" by Alexie Evan in (33)—originally not part of the corpus—exemplifies nearly all of the markers discussed above.

(33) (*Raven and His Two Wives*)

1 *Chulyin* *gun* *łu* *nutihna* *ve'uqa* *qiłan* *yen* *ki.*
Raven DEM NARR two wives there.were he too
'That Raven had two wives, that one.'

2 *Shan ghu k'ehzełgha.*
summer then they.were.drying.fish
'They were putting up fish in the summer time.'

3 *Shanlaghi uqu qel'an hnuyu shan quzeł.*
salmon they.were.fishing.for when summer they.spent
'They were fishing for salmon and that's how they spent the summer.'

4 *Iqech' t'qet'an k'ehzełgha k'ehzełgha.*
thus they.were.doing they.were.drying.fish they.were.drying.fish
'They kept on that way, putting up fish, drying it.'

5 *Q'uyehdi łiq'a hva łiq'a hva qisil* ***idi eł***
then salmon salmon were.gone CONJ

q'uyehdi łu dehi t'ihehghilyu ***ha'***.
then NARR cache they.put.them.in CONJ
'And then the salmon, the salmon stopped running and then they put them in the cache.'

6 *Vava nihdi qighishin nihyenilu* ***ha'***.
dry.fish stuff nicely they.put.it.away CONJ
'They carefully stored away the dry fish and everything.'

7 *Q'uyehdi hey qayeh ghu qech' nuhtasdatl'.*
then winter village there to.it they.took.off.again
'Then they left to return to their winter village.'

[3 lines omitted]

8 *Aa', iqech' t'qet'an ghu naqeli ghu łu hdalts'i*
yes thus they.did there fall there NARR they.stayed

hnuyu łu.
when NARR
'Ah, they kept doing that while they spent the fall there.'

9 *Q'uyehdi...*
then
'Then...'

10	***Q'u***	*hey*	***idi ełа***	*Chulyin*	*gun*	*t'etni*	*łu:*
	ADV	winter	CONJ	Raven	DEM	he.spoke.thus	NARR

'And now in winter, Raven spoke thus to them:'

11	*Shi*	*kitsa*	*shughu*	*tgheshyuni*	*vava*	*ghini*	*nda hqugh*
	1SG	next.time	EVID	I.want.to.go	dry.fish	DEM	how.much

shuq'u	*t'iłkugh.*
EVID	it.is.that.big

'"Let me go next time; I wonder how much dry fish is there?"'

Lines 1-4 are the general introduction: the story is set (in the summer), the main referents (Raven and his wives) are introduced, and it is explained what they are doing (drying fish). Lines 5-7 describe how they store the fish in a cache. While typically this is part of the routine of putting up fish, it is important for this story that this action be described in some detail—later in the story, Raven will go to the cache, try to pull out some fish, and die in the cache. For this reason, this part of the process is described very carefully.

Here, the discourse markers discussed above start appearing. The repeated occurrence of *q'uyehdi* 'then' marks the events as sequential: first the fish is being dried, then the salmon run stops, then everything is stored in the cache, then everyone returns to the winter village. Note that this sequentiality is complemented by the use of connectives in line final position, tying the lines closely together.

The omitted lines do not contain any discourse markers. They describe how Raven and his family stay in their winter camp and how his wives keep going to the cache to get fish. Line 8 above is the last line in the introduction of the narrative. Line 9, the isolated *q'uyehdi*, is right in the middle between the introductory paragraph and the rest of the narrative.

The beginning of the main part of the narrative is signaled by two discourse markers at the same time: *q'u* 'now', situating the narrative at a particular time (winter) and *idi ełа* 'and (suddenly)'. From its position in the utterance, we can assume that the meaning of *idi ełа* is 'suddenly' rather than a plain connective; it emphasizes how unusual it is that Raven wants to go to the cache himself, rather than sending his wives.

This example illustrates the functions of the discourse markers considered here:

- Connectives like *ch'u, idi ełа,* and *ha'* tend to create cohesion within paragraphs if they occur in line final position. In line initial position, they mark paragraph breaks.
- *Q'u* serves as a paragraph-defining marker. It frequently occurs in contexts where background information (such as the introduction to a narrative) is contrasted with foreground information (the main story line). Similarly, it can separate narrator comment or digressions from the main story line, and it frequently occurs when new action starts.
- Finally, the topic marker *-hdi* (no occurrence in the final examples) can signal the beginning of a new paragraph. Just as topic continuity is one of the main

> cohesion criteria within paragraphs, topic discontinuity (formally marked by *-hdi*) indicates a paragraph break.

This paper has shown how several Dena'ina discourse markers— *-hdi* 'new topic', *q'u* 'now', *ch'u, idi eła,* and *ha'* 'and'—can be used to structure Dena'ina traditional narratives. All of these particles have discourse functions, and these can only be grasped by the analysis of natural discourse.

While elicitation is a useful tool in determining certain kinds of semantic differences (for example the difference between *idi eła* 'and suddenly' vs. *ch'u* etc. and *ha/ha'* 'and'), it is impossible to understand the function of some of the discourse markers without considering their status in connected discourse. The glosses given by speakers for *-hdi* 'as for, how about' or *q'u* 'now' are fairly accurate, but not sufficient for linguistic inquiry. The study of texts reveals that *-hdi* and *q'u* function as new topic and new action markers, respectively, and it also reveals that the various conjunctions for 'and' have different scope, depending on their position within a line. An adequate description of these particles has to include information gleaned *both* from elicitation and from the study of connected speech.

Several discourse markers have not been addressed here for reasons of space, and also because their function is not fully understood. The corpus consists mainly of traditional narratives, and markers may function differently in other text genres. Also, it must be said that all or any or none of the discourse markers I have discussed may be used to define paragraphs in any particular instance; there is much that we do not know about Dena'ina discourse.

Appendix I

List of narratives. The dialect abbreviations are as follows: I–Inland dialect, O–Outer Inlet dialect, U–Upper Inlet dialect.

English Name	Dialect	Storyteller	Transcriber	Lines
Porcupine and Brown Bear	I	Gulia Delkittie	Tenenbaum	70
Ch'iduchuq'a Story	I	Alexie Evan	Tenenbaum	103
Two Women Story	I	Antone Evan	Tenenbaum	94
Ground Squirrel Story	I	Antone Evan	Tenenbaum	80
Raven and His Two Wives	I	Alexie Evan	Tenenbaum	62
Mouse Story	I	Mary Hobson	Lovick	90
Raven Story II	I	Mary Hobson	Lovick	31
Geese Story	I	Vonga Bobby	Kari	45
Grayling Story	I	Helen Dick	Kari	28
Caribou Story	I	Helen Dick	Berez	48
Raven Story I	O	Mary Oskolkoff	Kari	19
Underwater People Story	O	Fedosia Sacaloff	Kari	18
Crane Story	U	Katherine Nicolie	Kari	80
Quch' Nushjun Story	U	Shem Pete	Kari	182
Susitna story	U	Shem Pete	Kari	84
Hunting Dog Story	U	Shem Pete	Kari	132

References

Bobby, Vonga. 1991. Geese story. Recorded by Priscilla Russell Kari & Luther Hobson on 1 November, 1991. Transcribed by James Kari & Helen Dick. Recording TI3735, manuscript ANLC3735, Alaska Native Language Archive. Fairbanks, Alaska.

Chafe, Wallace. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (Advances in Discourse Processes 3). Norwood, NJ: Ablex Publishing Corporation.

Chafe, Wallace. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and grounding in discourse*, 21-51. Amsterdam & Philadelphia: John Benjamins.

de Reuse, Willem (see Reuse)

Delkittie, Gulia. 1974. Nini ggagga sukdu'a: Porcupine and brown bear story. Recorded by Joan Tenenbaum on 10 December, 1974. Recording TI1290, Alaska Native Language Archive. Fairbanks, Alaska.

Dena'ina Qenaga Digital Archive. http://qenaga.org (22 January, 2010.)

Dick, Helen. 2004. Grayling story. Recorded by Gary Holton on 4 June, 2004. Transcribed by James Kari. Recording TI4389, Alaska Native Language Archive. Fairbanks, Alaska. Manuscript courtesy the author.

Dick, Helen. 2005. Caribou story. Recorded by Andrea Berez & Gary Holton on 8 July, 2005. Transcribed by Andrea Berez & James Kari. Recording and manuscript courtesy the author.

Dryer, Matthew S. 1997. On the six-way word order typology. *Studies in Language* 21(1). 69-103.

Evan, Alexie. 1974a. Ch'iduchuq'a. Recorded by Joan Tenenbaum on 9 January, 1974. Transcribed by Joan Tenenbaum. Recording TI1285, Alaska Native Language Archive. Fairbanks, Alaska. Published in Tenenbaum 2006, 164-175.

Evan, Alexie. 1974b. Chulyin sukdu'a: Raven and his two wives. Recorded by Joan Tenenbaum in September 1974. Transcribed by Joan Tenenbaum. Recording TI1283, Alaska Native Language Archive. Fairbanks, Alaska. Published in Tenenbaum 2006, 102-109.

Evan, Antone. 1974a. Qunsha sukdu'a: Ground squirrel story. Recorded by Joan Tenenbaum on 20 November, 1974. Transcribed by Joan Tenenbaum. Recording TI1287, Alaska Native Language Archive. Fairbanks, Alaska. Published in Tenenbaum 2006, 134-141.

Evan, Antone. 1974b. Nutihna deghk'isna sukdu'a: Two women story. Recorded by Joan Tenenbaum on 20 November, 1974. Recording TI1287, Alaska Native Language Archive. Fairbanks, Alaska. Published in Tenenbaum 2006, 12-23.

Fernald, Theodore P. & Ellavina T. Perkins. 2006. Navajo coordination. In Andrea L. Berez, Suzanne Gessner, Leslie Saxon & Siri G. Tuttle (eds.), *Working Papers in Athabaskan Languages* (Alaska Native Language Center Working Papers 6), 91-132. Fairbanks: Alaska Native Language Center.

Givón, Talmy (ed.) 1983. *Topic continuity in discourse: A quantitative cross-language study* (Typological Studies in Language 3). Amsterdam & Philadelphia: John Benjamins.

Hobson, Mary. 2005. Chulyin sukdu'a: Raven story. Recorded by Olga Müller on 10 November, 2005. Recording TI4442, Alaska Native Language Archive. Fairbanks, Alaska. Transcript in Lovick 2007.

Hobson, Mary. 2006. Dlin'a sukdu'a: Mouse story. Recorded by Olga Lovick & Siri G. Tuttle on 6 June, 2006. Recording TI4499, Alaska Native Language Archive. Fairbanks, Alaska. Transcript in Lovick 2007.

Holton, Gary. 2004. Clause combining in Tanacross Athabascan. In Gary Holton & Siri G. Tuttle (eds.), *Working Papers in Athabaskan Languages* (Alaska Native Language Center Working Papers 4), 115-133. Fairbanks: Alaska Native Language Center.

Holton, Gary. 2005. Pitch, tone and intonation in Tanacross. In Sharon Hargus & Keren Rice (eds.), *Athabaskan prosody,* 249-275. Amsterdam: John Benjamins.

Holton, Gary & Olga Müller. 2005. Evidentiality in Dena'ina Athabascan. Paper presented at the Society for the Study of the Indigenous Languages of the Americas Annual Meeting. Oakland.

Holton, Gary & Olga Lovick. 2008. Evidentiality in Dena'ina. *Anthropological Linguistics* 50. 292-323.

Kalifornski, Peter. 1991. *K'tl'egh'i sukdu: A Dena'ina legacy: The collected writings of Peter Kalifornsky.* Fairbanks: Alaska Native Language Center.

Kari, James. 1975. A classification of Tanaina dialects. *Anthropological Papers of the University of Alaska* 17. 49-55.

Kari, James. n.d. Dena'ina stem file. Unpublished Lexware file.

Kari, James & James A. Fall. 2003. *Shem Pete's Alaska. The territory of the Upper Cook Inlet Dena'ina*, 2nd edn. Fairbanks: University of Alaska Press.

Kibrik, Andrej A. 2004. Coordination in Upper Kuskokwim Athabaskan. In Martin Haspelmath (ed.), *Coordinating constructions*, 537-553. Amsterdam: John Benjamins.

Lovick, Olga. 2006. *Agentivity and participant marking in Dena'ina Athabascan: A text-based study.* Cologne, Germany: University of Cologne dissertation.

Lovick, Olga. 2007. Dena'ina texts. Manuscript TI004L2007, Alaska Native Language Archive. Fairbanks, Alaska. Fairbanks, Alaska.

McCreedy, Lynn. 1989. Cohesion and discourse structure in three genres of Navajo discourse. In Eung-Do Cook & Keren Rice (eds.), *Athapaskan linguistics: Current perspectives on a language family* (Trends in Linguistics: State-of-the-Arts Report 15), 439-486. Berlin & New York: Mouton de Gruyter.

Mithun, Marianne. 2008. The extension of dependency beyond the sentence. *Language* 83. 69-119.

Nicolie, Katherine. 1976a. Ndał tsukdu: Crane story. Recorded by James Kari in 1976. Recording TI1270, Alaska Native Language Archive. Fairbanks, Alaska.

Nicolie, Katherine. 1976b. *Ndał tsukdu, gheldzay tsukdu: Crane story, moon story, edited by James Kari.* Fairbanks: Alaska Native Language Center.

Oskolkoff, Mary. 1976. Ggugguyni sukdu'a: Raven story. Recorded by Jim Kari on 30 March, 1976.

Pete, Shem. 1974. Susitnu htsukdu'a: The story of Susitna. Recorded and transcribed by James Kari on 10 or 11 October, 1974. Recording TI1270, Alaska Native Language Archive. Fairbanks, Alaska.

Pete, Shem. 1976. Quch' nushjun. Recorded and transcribed by James Kari in summer 1976. Recording TI1275, manuscript ANLC1275, Alaska Native Language Archive. Fairbanks, Alaska.

Pete, Shem. 1979. The hunting dog story. Recorded by James Fall on 8 February, 1979. Transcribed by James Kari. Recording TI3685, Alaska Native Language Archive. Fairbanks, Alaska. Manuscript published in Pete 1989. 3-7.

Pete, Shem. 1989. The hunting dog: A Dena'ina tale of subsistence values. *Alaska Fish and Game* 21(6). 3-7.

Reuse, Willem de. 2003. Evidentiality in Western Apache. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Studies in evidentiality*, 79-100. Amsterdam: John Benjamins.

Reuse, Willem de & Kevin Mullin. 2005. Western Apache recipes: Pragmatics and discourse structure. Paper presented at the Society for the Study of the Indigenous Languages of the Americas Annual Meeting. Oakland.

Rice, Keren. 1989. *A grammar of Slave*. Berlin: Walter de Gruyter.

Sacaloff, Fedosia. 1975. Unhsah tahna'ina: The underwater people. Published in Kari & Fall 2003. 356.

Saxon, Leslie. 1993. A personal use for Athapaskan 'impersonal' *ts'e-*. *International Journal of American Linguistics* 59(3). 342-54.

Saxon, Leslie. 1998. Complement clauses in Dogrib. In Leanne Hinton & Pamela Munro (eds.), *Studies in American Indian languages: Description and theory* (University of California Publications in Linguistics 131), 204-211. Berkeley: University of California Press.

Schiffrin, Deborah. 1987. *Discourse markers* (Studies in Interactional Sociolinguistics 5). Cambridge: Cambridge University Press.

Tenenbaum, Joan M. 1973. Field books A-M. Manuscripts TI973T1973f.1-TI973T1973f.14, Alaska Native Language Archive. Fairbanks, Alaska.

Tenenbaum, Joan M. 1978. *Morphology and semantics of the Dena'ina verb*. New York, NY: Columbia University dissertation.

Tenenbaum, Joan M. (compiler). 2006. *Dena'ina Sukdu'a: Traditional stories of the Tanaina Athabaskans.* Fairbanks: Alaska Native Language Center.

Thompson, Chad. 1989. *Voice and obviation in Athabaskan and other languages.* Eugene, OR: University of Oregon dissertation.

Tomlin, Russell S. 1997. Discourse semantics. In Teun A. van Dijk (ed.), *Discourse as structure and process* (Discourse Studies: A Multidisciplinary Introduction, vol. 1), 62-111. London: SAGE Publications.

Tuttle, Siri G. & Olga Lovick. 2007. Intonational marking of discourse units in two Dena'ina narratives. *Nouveaux Cahiers de Linguistique Française* 28. 305-316.

Wassillie, Albert. 1980. *Nuvendaltun ht'ana sukdu'a: Nondalton people's stories*. Fairbanks: Alaska Native Language Center.

Olga Charlotte Lovick
olga@lithophile.com

9

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 203-222
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/

Be careful what you throw out: Gemination and tonal feet in Weledeh Dogrib

Alessandro Jaker

Stanford University

The Weledeh dialect of Dogrib (Tłı̨chǫ Yatìì) is spoken by people of the Yellowknives Dene First Nation, in and around Yellowknife, Northwest Territories. Within the formal framework of Lexical Phonology (Kiparsky 1982), this paper argues for an over-arching generalization in the phonology of Weledeh Dogrib: the constraint NoContour-Ft, which prefers (High-High) and (Low-Low) feet, but militates against (High-Low) and (Low-High) feet. NoContour-Ft is satisfied differently in different morphophonological domains: vowel deletion at the Stem Level, gemination at the Word Level, and High to Mid tone lowering at the Postlexical Level. This analysis requires that consonant length be treated as phonological in Dogrib—that is, consonant length contributes to syllable weight and mora count—even though there are no minimal pairs based on consonant length. Similarly, the distinction between High and Middle tone does not distinguish any lexical items, but is nevertheless important for the prosody of the language. Thus the paper makes a methodological point about the importance of allophonic alternations for phonological theory. Our view of what counts as contrastive or allophonic, however, is to a large extent theory-dependent; therefore, the paper also emphasizes the importance of phonetic measurements when doing fieldwork.

1. INTRODUCTION: WHAT COUNTS AS PHONOLOGICAL?

> ...even such commonplace categories as subject and verb are theoretical constructs, which may or may not be the ones most appropriate for the data under consideration. (Gil 2001: 126)

The most basic task in fieldwork is simply to transcribe what we hear—whether live or from a recording, we take a noisy and chaotic speech signal, and reduce it to an idealized set of characters on the page. Without imposing some kind of order on the data, it would be impossible to talk about higher levels of grammar—subject agreement, noun incorporation, and so forth. At the same time, how we construct phonological categories depends on the methodology and assumptions of the fieldworker just as much as the speech signal itself. Our assumptions are often not our own, but those inherited from previous generations of linguists—Athabaskanists, in particular, have the benefit of a descriptive tradition going back over 100 years (Goddard 1912; Li 1933, 1946). When I talk to other linguists about geminate consonants in Dogrib, for example, I am often asked, with an air of skepticism, "are there minimal pairs?" For the American structuralist, the reason for this question is obvious: whether or not we should talk about phonological consonant length depends

on whether consonant length can distinguish utterances (Bloomfield 1933). Structuralist descriptions of phoneme inventories have been left largely intact by generative fieldworkers, even when the assumptions upon which these descriptions are based would no longer be considered valid in a generative context. In this article, I will re-examine some basic assumptions about the phonology of Weledeh Dogrib, a short summary of which is given in (1).

(1) *Generally accepted views about Dogrib:*
   i. The final syllable is stressed (Marinakis 2004).
   ii. Coda consonants in Dogrib are not moraic (Marinakis 2004).
   iii. Consonant length in Athabaskan languages is phonetic, not phonological (McDonough and Ladefoged 1993, Tuttle 2005).

What we shall see is that even these simple descriptive statements, though perfectly well-motivated in earlier frameworks, are misleading when taken out of their original context and applied at face value in a new theory—in my case, that of Lexical Phonology (Kiparsky 1982) and Optimality Theory (OT). For example, it makes perfect sense for a Prague School phonologist to say that the final syllable, also the stem syllable, is stressed in Dogrib, since stress is understood as a position which licenses more contrast (Trubetzkoy 1939), and indeed there is a larger inventory of segments allowed in the stem than in the prefixes (Rice 1989, Marinakis 2004). In OT, however, we more closely associate stress with some set of phonetic properties, such as increased F0, amplitude, and duration—in which case this claim turns out to be false (in many cases). To say that consonant length is "not phonological" in an American structuralist framework means only that consonant length cannot distinguish utterances, which is largely true; yet in OT, this means that the phonetic length of consonants does not affect the number of moras, which in turn affects syllable weight. As I will show, the key generalizations about morphophonemics in Dogrib require reference to syllable weight, which in turn requires reference to consonant length—we would miss important generalizations if geminates were not included as part of the phonology.

For me, then, fieldwork on Dogrib has been a process of unraveling layers of unstated assumptions, both others' and my own. Relying on descriptive statements made by others means adopting their assumptions about what facts ought to be included in the description and what should be thrown out. Fieldwork with speakers of Weledeh Dogrib enables me to go back to the original speech signal and decide for myself what is structurally important and what is not. As I will demonstrate, once consonant length is included as part of the phonological representation for Weledeh Dogrib, a whole new set of descriptive generalizations emerge, the most important of which is the existence of the *tonal foot*, a unit of suprasegmental structure that simultaneously regulates tone, stress, and syllable weight (section 2).

Before proceeding further, a note on the language. The Weledeh dialect of Dogrib (Tłı̨chǫ Yatiì) is a northern Athabaskan language, spoken by people of the Yellowknives Dene First Nation, in the communities of Dettah and Ndilǫ, near Yellowknife, Northwest Territories. While other dialects of Dogrib are still being acquired by children, due to the community's proximity to Yellowknife, the Weledeh dialect has very few fluent speakers

under age 40. The community is traditionally bilingual: many older people also speak Chipewyan (Dëne Sųłiné) as well as Dogrib. The Goyatikǫ̀ Language Center in Dettah is actively working to preserve and revitalize both traditional languages through translator and interpreter training, literacy training, recording elders' stories, and making teaching materials for children. The data in this paper are taken from six weeks of my own fieldwork in Yellowknife in the summer of 2005. My primary elicitation was with two speakers, Mary Louise Drygeese and Michel Paper, and checked with a third speaker, Mary Rose Sundberg, for accuracy in transcription. The phonetic data presented here were digitally recorded on a Marantz CDR-300 recorder and analyzed in Praat (Boersma & Weenik 2007).[1]

2. TONAL FEET. My central thesis is that several seemingly unrelated morphophonological processes in Weledeh Dogrib can be explained by reference to a unit of representation called the *tonal foot*. A tonal foot is a unit of metrical structure which is sensitive to tone. Specifically, tonal feet are required to be *level*: both syllables in a disyllabic foot are required to have the same tone. While there is no consensus in the literature on how to define 'tonal foot', and several versions have been proposed (Rice 1990, Zec 1999, De Lacy 2002), they all have in common some form of interaction between tone and metrical structure. In Dogrib specifically, I claim that the tonal foot is a moraic trochee which is subject to a restriction that there be no contour tones within the foot. That is, the canonical disyllabic foot in Dogrib consists of two light syllables in which the first syllable is stressed, and both have the same tone, i.e. either (High-High)[2] or (Low-Low). Similarly, monosyllabic feet should consist of a heavy syllable and not contain a contour tone. A typology of acceptable and unacceptable tonal foot types in Dogrib is given in Table 1.

[1] Many thanks to Mary Rose Sundberg and Betty Harnum of the Goyatikö Language Center in Dettah, Northwest Territories, and the people of the Yellowknives Dene First Nation, for all their support, both practical and moral. Thanks also to Mary Louise Drygeese and Michel Paper for the many hours they spent working with me and answering my questions. Thanks also to Will Leben, Paul Kiparsky, Arto Anttila, Larry Hyman, Rebecca Scarborough, Kevin Ryan, Marianne Mithun, Spike Gildea, Andrea Berez, participants at the 2007 WAIL conference, and one anonymous reviewer for comments on previous versions of this article. All remaining errors are my own.

[2] I follow the orthographic convention of leaving High tone unmarked in Dogrib (*a*), while Low tone is marked with a grave accent (*à*).

TABLE 1. Typology of tonal feet

Tonal foot type	Good or bad	Example
a. High-High trochee H (σ σ)	good	L H L *dze*(*k'oo*)(*lane*) 'wild rose'
b. Low-Low trochee L (σ σ)	good	L L L *bo*(*kawi*)*t'e* '2 of us will cook'
c. High-Low trochee H L (σ σ)	bad	L H L H *(*boka*)(*whewi*)*t'e* '2 of us have cooked'
d. Low-High trochee L H (σ σ)	bad	L H H *(*shets'e*)*zhe* 'we eat'
e. Low-Mid trochee L M (σ σ)	acceptable as last resort	L M H (*shets'e*)*zhe* 'we eat'

The examples in (a) and (b) illustrate well-formed tonal feet. The examples in (c) and (d) are forms which we would expect to surface in Dogrib if there were no tonal restriction on foot form, that is, they represent a faithful parse of their respective underlying forms. Candidate (e), a Low-Mid trochee, is acceptable as a last resort to repair a Low-High trochee, if no other repair strategy can be used. In the remainder of this paper, I will illustrate how several phonological processes in Dogrib are motivated by tonal feet, relying crucially on the constraint NOCONTOUR-FT (Pearce 2006), which requires that feet be level. Within the framework of Lexical Phonology (Kiparsky 1982, 2000; Mohanan 1986; Hargus 1988), I consider how the tonal foot can provide a unified analysis of phonological processes in Dogrib at the Stem, Word, and Postlexical levels.

3. DERIVING TONAL FEET. In Table 2, I give the surface footing for the imperfective and perfective paradigms of the verb *bòkà√t'è* 'cook', in Dogrib. As these feet are metrical feet, the strong position of each foot receives at least a secondary stress, and the strong posi-

tion of the rightmost foot receives primary lexical stress. As we would expect in a moraic trochee language (Hayes 1995), all of the feet in Table 2 contain either two light syllables or a single heavy syllable.[3] These facts are, on the surface, consistent with ordinary moraic trochees, without any reference to tone. The evidence that tone also plays a role in the metrical structure of Dogrib comes from morphophonemics: in some cases, the surface forms, as shown in Table 2, have been modified significantly from their underlying forms, to conform to a certain prosodic pattern. Specifically, as I will argue, phonological processes in the language conspire to avoid High-Low or Low-High sequences within a foot.

TABLE 2. Paradigm for bòkà√t'è 'cook', surface forms, with footing

Imperfective[4]

	Singular	Dual	Plural
1st person	(*bòkà*)(*eh*)*t'è*	*bò*(*kàì*)*t'è*	(*bòkà*)(*ts'ee*)*t'è*
2nd person	*bò*(*kàį*)*t'è*	*bò*(*kàah*)*t'è*	*bò*(*kàah*)*t'è*
3rd person	(*bòkà*)(*et*)*t'è*	(*bòkà*)(*gee*)*t'è*	(*bòkà*)(*gee*)*t'è*

Perfective

	Singular	Dual	Plural
1st person	(*bòkà*)(*whih*)*t'e*	*bò*(*kàwhì*)*t'e*	(*bòkà*)(*ts'ih*)*t'e*
2nd person	(*bòkà*)(*whęę*)*t'e*	(*bòkà*)(*whah*)*t'e*	(*bòkà*)(*whah*)*t'e*
3rd person	(*bòkà*)(*whet*)*t'e*	(*bòkà*)(*geh*)*t'e*	(*bòkà*)(*geh*)*t'e*

In order to characterize tonal feet more precisely, let us start with the simplest case, the case in which the input string already conforms to the tonal restrictions of the language, and so can be parsed as a metrically well-formed sequence as is. An example of this is shown in Table 3 below. In Table 3, the input /bòkàgeet'è/ is parsed into moraic trochees, using three constraints that are standard in the literature: FOOTBINARITY, PARSESYLLABLE, and ALIGNRIGHT.

[3] I assume that a syllable can be made heavy either by having a long vowel (CVV) or having a coda consonant (CVC), which includes geminates.

[4] In (*bòkà*)(*eh*)*t'è* 'I cook' the sequence *àe* is pronounced as two syllables and in *bò*(*kàį*)*t'è* 'you (SG) cook' the sequence *àį* is pronounced as a single syllable.

TABLE 3. Feet assigned atonally

/bò-kà-ge-e-t'è/	FTBIN	PARSE(σ)	ALIGN-R(FT, PRWD)
a. *bò*(*kà.gee*)*t'è*	*!	**	*
b. *bòkà*(*gee*)*t'è*		***!	*
☞c. (*bò.kà*)(*gee*)*t'è*		*	***

FOOTBINARITY refers to a general property of moraic trochees: they must have exactly two moras, either from a single heavy syllable, or from two light syllables. Candidate (a) is a violation of FOOTBINARITY because it contains the foot (*kà.gee*), which consists of a light syllable followed by a heavy syllable, giving three moras. The constraints PARSE(σ) and ALIGN-R(FT, PRWD) refer to the way in which feet are constructed within the word. The constraint PARSE(σ) penalizes any syllable that is not part of some foot. The constraint ALIGN-R(FT, PRWD), on the other hand says that every foot should be aligned with the right edge of some prosodic word. If this constraint were undominated, there would only be a single foot at the very right edge of each word. Because ALIGN-R(FT, PRWD) is dominated by PARSE(σ), however, additional feet are created, as in candidate (c), albeit as far to the right as possible.

Another constraint not considered in Table 3 is NONFINALITY. NONFINALITY requires that the final syllable of a prosodic word not be footed. The interaction of NONFINALITY with FOOTBINARITY and PARSE(σ) is shown in Table 4.

TABLE 4. NONFINALITY in metrical parsing

/bò-kà-whe-ne-t'e/	FTBIN	NONFINALITY	PARSE(σ)
a. (*bò.kà*)(*whe.ne.t'e*)	*!	*	
b. *bò*(*kà.whe*)(*ne.t'e*)		*!	*
☞c. (*bò.kà*)(*whe.ne.*)*t'e*			*

Table 4 illustrates the parsing of the form *bòkàwhenet'e*, which is an attested variant of *bòkàwhęęt'e*, shown in Table 2. Candidate (a), even though it perfectly satisfies PARSE(σ) by including every syllable in some foot, contains a tri-moraic foot (*whe.ne.t'e*) and violates NONFINALITY, as does candidate (b). Candidate (c) is therefore optimal, in that it satisfies FOOTBINARITY and NONFINALITY, with only one violation of PARSE(σ). Finally, one relevant constraint not shown in Tables 3 and 4 is RHYTHMTYPE=TROCHEE, which ensures that feet in Dogrib are trochaic (i.e. strong-weak) rather than iambic (weak-strong).

With regard to stress, the foot parsing shown in candidate (c) predicts that there should be stresses on both *bò* and *whe* since these are in the strong position of trochaic feet. We would expect these syllables to show the usual phonetic correlates of stress. Thus, a stressed, short vowel with lexical Low tone, such as *bò*, would be expected to exhibit longer duration and a higher F0, but not so long as to encroach on a lexically long vowel,

nor as high to encroach on a lexically High tone, in accordance with contrast and dispersion (Flemming 1996, 2001). Future instrumental studies will be necessary to show if this is correct.

Now, given that something like the constraint-ranking shown in Tables 3 and 4 can successfully parse all of the words shown in Table 2, why should the metrical phonology of Dogrib make any reference to tone at all? What is *tonal* about tonal feet?

Table 5 presents five different inputs, and contrasts the result we would expect with standard right-to-left moraic trochees (Hayes 1995) versus the actual result, derived from tonal feet. In (a), there is no difference between simple moraic trochees and tonal feet, because the input string happens to already conform to the restriction that feet be level. In (b), however, we see that the stem-initial consonant geminates, so that (*whet*) is made to form a High-toned foot by itself, while (*bòkà*) forms a Low-toned foot to its left. In (c), (*bò.kà*)(*whe.ne*)*t'e* is one of the attested possibilities in Weledeh Dogrib, but so are the forms (*bò.kà*)(*whęę*)*t'e* and (*bò.kà*)(*whįį*)*t'e*, which have undergone a process which I call *nasal coalescence*, the latter of these having also undergone *nasal raising*. While these processes do not make direct reference to tone, they are sensitive to feet, in that they occur only within but not across foot boundaries. In (d), we see vowel syncope, perhaps the most dramatic process conditioned by tonal feet. In (d), the vowel of the prefix /whe/, which has underlying High tone, is deleted, since there is no way to parse it into a level foot. Based on examples such as these, I will use the constraint NoContour-Ft (Pearce 2006) to formally characterize the way non-level feet are penalized in Dogrib. Finally, (e) represents a prima facie counterexample to the existence of tonal feet, since the actual form, *bò.*(*kàį*)*t'è*, contains the non-level foot (*kàį*), while the form which we would expect based on standard moraic trochees does not. In this case, it is necessary to say that the constraint NoContour-Ft is dominated by other constraints, in particular Align-R(Ft, PrWd), which seeks to reduce the overall number of feet in a Prosodic Word, even if some of the resulting feet are non-level.

Table 5. Footing with a-tonal and tonal feet

Input	Expected Result (standard moraic trochees)	Actual Result (with tonal feet)
a. /bò-kà-ge-e-t'è/	(*bò.kà*)(*gee*)*t'è*	(*bò.kà*)(*gee*)*t'è*
b. /bò-kà-whe-t'e/	*bò.*(*kà.whe*)*t'e*	(*bò.kà*)(*whet*)*t'e*
c. /bò-kà-whe-ne-t'e/	(*bò.kà*)(*whe.ne*)*t'e*	(*bò.kà*)(*whe.ne*)*t'e* or (*bò.kà*)(*whęę*)*t'e* or (*bò.kà*)(*whįį*)*t'e*
d. /bò-kà-whe-wìd-t'e/	(*bò.kà*)(*whewì*)*t'e*	*bò.*(*kàwhì*)*t'e*
e. /bò-kà-e-ne-t'è/	(*bò.kà*)(*e.ne*)*t'è*	*bò.*(*kàį*)*t'è*

4. GEMINATION. While the existence of phonetically long consonants in some Athabaskan languages has been previously noted in the literature (McDonough & Ladefoged 1993, Tuttle 2005), previous works on Dogrib (Coleman 1976, Ackroyd 1982, Marinakis 2004) make no mention of geminates. In arguing for phonological geminates in Weledeh Dogrib, I will present both phonetic evidence for a categorical distinction between singleton and geminate consonants, as well as a phonological account of their distribution.

Phonetically, it seems that geminate consonants in Dogrib are 1.7-2.0 times the length of their singleton counterparts on average. This is consistent with what is known about the phonetic realization of geminates cross-linguistically (Keer 1999). The durations for singleton and geminate /t'/, /n/, and /l/ are shown in Figure 1 below.

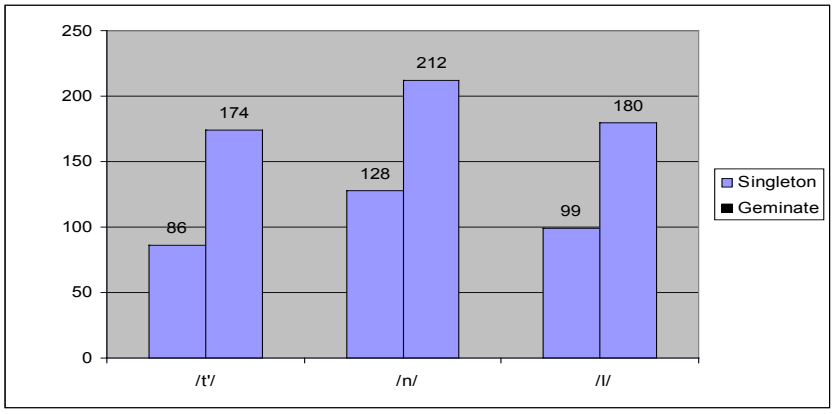


FIGURE 1. Singleton and geminate durations for /t'/, /n/, and /l/

The data in Figure 1 are based on a relatively small number of tokens: approximately 40 tokens each for /n/ and /l/, and only 12 tokens for /t'/. Nevertheless, in all cases, the difference between singleton and geminate consonants was categorical and highly significant, $p<0.001$ (*t*-test assuming equal variances). Thus when I refer to 'geminates' in the remainder of this article, I am referring specifically to consonants whose duration is 174-212 ms on average, while by 'singleton' I mean consonants that are 86-128 ms on average, as shown in Figure 1 above.

Given these facts, let us pause to consider whether consonant length in Dogrib ought to be considered 'phonetic' or 'phonological'. Under American structuralism, some category is phonological if it is able to distinguish utterances (Bloomfield 1933). Consonant length in Dogrib seems to fail this test, as there are few if any minimal pairs. On the other hand, in generative phonology, including OT, phonological status is a representational issue, not a question of contrastiveness *per se*. For example, in Italian, consonant length is contrastive, while vowel length is conditioned by stress. In the form *vá:do* 'I go,' the intial vowel is lengthened because it is stressed and in a penultimate, open syllable. There could be no minimal pair with **vádo* (short vowel) because a long vowel is required by the grammar. On the other hand, if one considers a somewhat more abstract level of representation—

namely, foot structure—(*vá:*)*do* is entirely parallel to (*ván*)*no* 'they go'. In both cases, the penultimate syllable forms a moraic trochee by itself: the former has two moras linked to a vowel, the latter has one mora linked to a vowel and another to the geminate consonant. If vowel length were removed from a phonological description of Italian, no lexical contrasts would be merged, yet at the same time important generalizations would be lost. We would not know which syllables count as light or heavy, which would in turn change our predictions about stress and segmental processes conditioned by stress.

Dogrib represents, in some sense, the mirror-image of Italian: vowel length is contrastive in Dogrib, while consonant length is prosodically conditioned. When I first began fieldwork with speakers of Weledeh Dogrib, I started hearing geminate consonants almost immediately. At first, I refused to believe my own ears: "this is just my bias as a native speaker of Italian," I thought. "Just another eurocentric category I'm trying to impose." What convinced me ultimately was that it was necessary to talk about geminates if the prosodic system was going to make any sense. I accepted geminates when I began to understand the language in terms of its own patterns. The pattern—in this case the tonal foot—is a rather abstract and unusual one. But I would never have discovered it if I had ignored geminates in the first place. Sometimes, then, a phonemic analysis is a self-fulfilling prophecy: by excluding some category from the system, we destroy the patterns that provide evidence for that category's importance. It is only by going back to the original speech signal that we can recover patterns that may have been overlooked, and the phonetic evidence for those patterns.

In my analysis, geminates are conditioned by tonal feet in the sense that the constraint NoContour-Ft plays a crucial role. This is shown in Table 6.

Table 6. Analysis of gemination in Dogrib (word level)

/bò-kà-(e)t'è/	FtBin	DepAssoc (V, μ)	Max(V)	NoCntr-Ft	DepAssoc (C, μ)
a. (*bòkà*)(*e*)*t'è*	*!				
b. (*bòkà*)(*ee*)*t'è*		*!			
c. (*bòkà*)*t'è*			*!		
d. *bò*(*kà.e*)*t'è*				*!	
☞e. (*bòkà*)(*et*)*t'è*					*

In the form *bòkàett'è*, *e* is a conjunct prefix while *bò* and *kà* are disjunct prefixes. Since I assume that conjunct prefixes are added at the Stem Level, *e* is shown as already footed in the input to the Word Level phonology, although this is not crucial to the present example.

In Table 6, each of the candidates (a)-(d) is ill-formed in the language because they violate some high-ranking constraint. Candidate (a) violates FootBinarity, on account of (*e*), which is a foot consisting of a single, light syllable. Candidate (b) violates a constraint against vowel lengthening, formulated here as DepAssoc(V, μ) (see section 5 for discus-

sion), while candidate (c) violates a constraint against vowel deletion. Finally, candidate (d) is ill-formed because it contains a Low-High trochee, which violates NoContour-Ft.

It is the ill-formedness of candidates such as (d) in Table 6 that provide evidence for tonal feet. That is, (d) is what we would expect to surface as optimal in an ordinary moraic trochee language, but it is not the optimal candidate in Dogrib, since it contains a Low-High trochee. Instead, the foot boundary is shifted one syllable farther to the left, and the winning candidate contains the Low-Low trochee *bòkà*. This means that the penultimate syllable must now form a foot all by itself, which means, in turn that weight must be added to it, otherwise it would be a degenerate (i.e. monomoraic) foot, as in candidate (a). There are two ways this could be accomplished: either lengthen the vowel itself, or geminate the following consonant. In Dogrib, vowel length is contrastive, which is expressed formally as a high-ranked constraint demanding faithfulness to underlying vowel length—DepAssoc(V, μ). On the other hand, DepAssoc(C, μ), which militates against consonant lengthening, is low-ranked, and so gemination is preferred over vowel lengthening, and thus candidate (e) emerges as the winner.

Here again it is important to stress the importance of allophonic processes in phonology. The constraints FootBinarity and NoContour-Ft, taken together, express the generalization that feet in Dogrib have two moras, both with the same tone. In a language where each affix carries its own tone and literally thousands of affix-combinations are possible, ill-formed sequences are bound to arise just by morpheme concatenation. So, how does Dogrib maintain level, bimoraic feet? That the constraint DepAssoc(V, μ) is high-ranked tells us two things: vowel length is contrastive, and there cannot be phonological processes that alter vowel length. Conversely, that DepAssoc(C, μ) is low-ranked tells us that consonant length is not contrastive and there *can* be processes that alter consonant length. In other words, the effects of tonal feet (or any other phonological restriction) are most visible on categories whose functional load is lowest. The surest way to miss generalizations, then, is to throw out allophonic processes. On the other hand, Americanist fieldwork is rooted in a tradition which, historically, threw these out as a methodological principle. The only way out of this problem, in my opinion, is to keep digital sound recordings of all the data one collects. Even the simplest descriptive statements are biased by my own assumptions, which may be just as misguided as those of previous decades. I can only hope to leave behind enough phonetic detail that a linguist 100 years from now will have no trouble refuting my analysis!

5. SYNCOPE. Perhaps the most dramatic evidence for tonal feet in Dogrib is vowel syncope. Vowel syncope is a process in which, in an underlying sequence of two light syllables, with a High-Low tone pattern, the High toned vowel deletes. This type of process is very unusual typologically, and has in fact been claimed to be impossible (Blumenfeld 2006). An example of this process is shown in Table 7 below.

Table 7 treats the Stem Level phonology, which, I assume, involves only the stem and the conjunct prefixes, /e/ and /wìd/. The disjunct prefixes /bò/ and /kà/ are shown in light gray, to indicate that they are still "invisible" to the phonology, as they will be affixed in a later cycle. The driving force behind the syncope process in Table 7 is the constraint NoContour-Ft, which militates against candidates such as (a), which faithfully preserve

a High-Low sequence from the input. That being the case, the other constraints in Table 7 serve to decide what repair strategy should be used.[5] The constraint which forces some sort of deletion is MaxAssoc(V, Tone), based on Myers 1997.

Table 7. Tone-conditioned vowel deletion (stem level)

/bò-kà-e-wìd-t'è/	NoCntr-Ft	MaxAssoc (V, Tone)	MaxAssoc (μ, V)	Max(L)	Max (μ)	Max (H)
a. *bò-kà-(ewì)t'è*	*!					
b. *bò-kà-(èwì)t'è*		*!				*
c. *bò-kà-(ewi)t'è*		*!		*		
d. *bò-kà-(wìì)t'è*			*!			*
e. *bò-kà-(ee)t'è*			*!	*		
f. *bò-kà-(e)t'è*				*!	*	
☞g. *bò-kà-(wì)t'è*					*	*

MaxAssoc(V, Tone)

(*informally*)*:* For every vowel associated with a tone in the input, it must remain associated with its tone in the output. For every output vowel not associated to its input tone, assign a violation mark.

Crucially, this constraint is violated by any output vowel which has been de-linked from its tone; therefore, it cannot be violated if there is no output vowel. In this way, the constraint MaxAssoc(V, Tone) is able to force vowel deletion: just as a captain goes down with the ship, a vowel must go down with its tone.

In a similar fashion, moras are also forbidden from de-linking from their input vowels.

MaxAssoc(μ, V)

(*informally*): For every mora associated with a vowel in the input, it must remain associated with its vowel in the output. For every output mora not associated to its input vowel, assign a violation mark.

The constraint MaxAssoc(μ, V) serves to rule out candidates such as (d) and (e) in Table 7, in which compensatory lengthening has taken place. This is because the nature of compensatory lengthening is that some segment deletes and its mora re-associates to

[5] Gemination is, in principle, another possible strategy, though in my analysis, it is available only at the Word Level. The constraint ranking at both the stem level and postlexical level forbids altering consonant length (though the relevant constraints are not shown in Figure 9).

another segment. That tones and moras should behave this way leads us to a rather curious conclusion: the autosegmental phonology of Dogrib doesn't seem very autosegmental at all, and tones and moras behave like properties of segments. Why should this be?

There is of course no a priori reason to expect that tones and moras should float around freely on their own tier, any more than place or manner of articulation. On the other hand, the most typical situation is for tones and moras to be *stable*, that is, to remain behind even after their segment deletes. Furthermore, if some sequence of tones is ill-formed in a language, generally the repair strategy will be to fix the tone sequence, e.g. by spreading or downstepping, and not alter anything on the segmental tier. Why should Dogrib be any different?

The answer, in my opinion, has to do with the relatively recent tonogenesis in Athabaskan languages (Krauss 2005). Before tonogenesis, vowels which in modern Dogrib have Low tone were followed by a glottal stop (ʔ), while modern High tones were not. This means that Low toned vowels necessarily formed closed syllables, while High tones were mostly open syllables. This means, in other words, that modern (High-Low) sequences were historically (light-heavy) sequences. This would suggest that, originally, syncope in Dogrib was based on stress and weight: the syncope rule targeted the weak branch of an iambic foot. With tonogenesis, the final glottals were lost, and with it was lost any transparent motivation for syncope in terms of weight. Instead, the syncope rule was re-interpreted as being based on tone, which gave rise to the constraint NoContour-Ft.

Finally, one should note that the winning candidate in Table 7 represents the output to the Stem Level phonology, and is not the final output. Specifically, the glide *w* in candidate (g) will be deleted as part of a process of vowel coalescence, to be shown in section 6.2, and thus the actual output is *bòkàìt'è*.

6. COALESCENCE PROCESSES. A majority of morphophonemic alternations in Dogrib involve some form of coalescence. Marinakis (2004) argues that coalescence is driven by constraints of the type *Struc(σ), which seek to reduce the overall number of syllables within the word. While it is true that coalescence processes almost invariably do reduce the overall number of syllables, this analysis makes relatively few predictions about which syllables should be the targets of coalescence. Here I propose instead how coalescence can be derived from tonal feet. The driving force here, I argue, is the Stress-to-Weight principle, or SWP (Borrelli 2000). This is a constraint which requires, if a syllable bears primary lexical stress, that it be heavy, i.e. have two moras. This constraint poses problems for a moraic trochee system, in that by simply lengthening stressed syllables, one incurs violations of foot binarity. A light-light trochee of the form (**μ** . μ), which bears primary lexical stress, satisfies FootBinarity but violates Stress-to-Weight, while, conversely, a heavy-light trochee of the form (**μ μ** . μ) satisfies Stress-to-Weight but violates FootBinarity.

There is a way around this problem, however. Recall that, in a moraic trochee language, a well-formed trochee can consist of either two light syllables (light-light) or a single heavy syllable (heavy). A solution, therefore, is to coalesce both syllables of a (light-light) trochee into a single, heavy syllable. The stressed syllable is then heavy, while at the same time the foot is still binary. In Dogrib, this is accomplished by deleting intervocalic consonants, and merging together the newly adjacent vowels. The consonants which delete

in Dogrib do not seem to form a natural class: *d, r, n,* and *w* delete intervocalically (under certain prosodic conditions), while *t, l, m,* and *wh* do not. While the seeming unnaturalness of this class of segments poses a theoretical problem, the issue is largely independent of the issue of tonal feet. Therefore, in the following sections, I will use the term 'sonorant' as a placeholder for these segments, and make use of the constraint MAX(son) to refer to the faithfulness violations incurred by deleting one of these segments.

6.1. NASAL COALESCENCE. Nasal coalescence is a process by which the consonant *n*, when situated between two short oral vowels, deletes, and leaves behind a single, long nasal vowel. This process is variable, and also interacts with another process called nasal raising, in which nasal *ą* is raised to *ǫ*, and *ę* is raised to *į*. Thus for the input /nà-whe-ne-zè/ 'you (sg) have hunted', *nàwhenezè*, *nàwhęęzè*, and *nàwhįįzè* are all possible outputs (*nàwhinizè*, however, is unattested). In Figure 8, below, I show how the grammar of Dogrib can derive both *bòkàwhenet'e* and *bòkàwhęęt'e*, two variants meaning 'you (sg) have cooked'.

TABLE 8. Nasal coalescence (stem level)

/ bò-kà -whe-ne-t'e/	MAX [±Nas]	FTBIN	SWP	MAX(son)
☞a. *bò-kà*-(*whęę*)*t'e*				*
(☞) b. *bò-kà*-(*whene*)*t'e*			*(!)	
c. *bò-kà*-(*whę*)*t'e*		*!	*	*
d. *bò-kà*-(*whee*)*t'e*	*!			*

The driving force behind the process of nasal coalescence in Table 8 is the constraint SWP, the Stress-to-Weight Principle, which requires that stressed syllables be heavy. In this case, however, the process itself is optional, and so candidate (b) is an attested form in Dogrib, albeit less preferred. In candidate (b), the foot (*whene*) consists of two light syllables, of which the first one is stressed. One could simply delete the nasal consonant, leaving a long oral vowel, as in candidate (d), although this violates the constraint MAX[±Nas], which demands that the nasal feature be preserved. Therefore, the preferred output is candidate (a), which contains a long nasal vowel.

6.2. COALESCENCE AND GLIDE DELETION. In a similar way, the glide *w* deletes intervocalically in order to create a heavy syllable. This is illustrated in Table 9.

TABLE 9. Illustration of w-deletion (word level)

/bòkà(wì)t'è/	FTBIN	DEP(μ)	SWP	MAX(son)
a. (*bòkà*)(*wì*)*t'è*	*!		*	
b. (*bòkà*)(*wìì*)*t'è*		*!		
c. *bò*.(*kàwì*)*t'è*			*!	
☞d. *bò*.(*kài*)*t'è*				*

Table 9 is a continuation of Table 7. The original input was /bò-kà-e-wìd-t'è/ 'two of us are cooking,' where the output of the Stem Level phonology is *bòkàwìt'è*, and the output of the Word Level is *bòkàìt'è*, which is the actual surface form shown in the paradigm in Table 2. As with the previous example, the driving force behind coalescence in Table 9 is the Stress-to-Weight principle. In candidate (c), *bò*(*kàwì*)*t'è*, main stress falls on *kà*, which is a light syllable. This is a violation of the SWP, and so candidate (d) is preferred, which deletes the intervocalic *w* and leaves behind the diphthong *ài*, creating a heavy syllable.

7. TONE LOWERING. Finally, we come to the phenomenon of High to Mid tone lowering. This is a postlexical process which is used as a last resort to repair (Low-High) trochees that cannot be repaired by other means. Phonetic evidence for this process is given in Table 14 below. First, however, it is necessary to address the question of how these ill-formed sequences arise in the first place. I propose that the grammar is forced to create (Low-High) trochees at the Word Level in trisyllabic words, in order to avoid having two geminates in a row. This is illustrated in Table 10 using the word *shèts'ezhe* 'we eat'.

TABLE 10. Creation of non-level feet in trisyllabic words (word level)

/shè-(ts'e)zhe/	FTBIN	[*C_{μ}]^{2}PRWD	PARSE(σ)	NOCONTOUR-FT	*Cμ
a. *shè*(*ts'e*)*zhe*	*!		**		
b. (*shèt*)(*ts'ez*)*zhe*		*!	*		**
c. *shè*(*ts'ez*)*zhe*			**!		*
☞d. (*shèts'e*)*zhe*			*	*	

Based on the grammar of Dogrib developed so far, we might expect something like candidate (b) to surface as optimal. In this candidate, gemination has happened twice, to create two heavy syllables, *shèt* and *ts'ez*, each of which forms its own foot. What rules it out is a constraint on two adjacent geminates, also known as Schneider's Law (Lipscomb 1992, Dresher & Johns 1995). The precise formulation of this law is problematic (see Dresher & Johns 1995 for discussion), so here as a sort of placeholder I have stated it as [*C_{μ}]^{2}PRWD, or "no two geminates within a prosodic word." This restriction is well-doc-

umented in Inuktitut, a language with which the Yellowknives Dene were in contact until the early 20th century. In short, gemination will not work as a repair strategy for trisyllabic Low-High-High words, and so some other repair strategy must be employed, which is to lower the first High tone to a Mid tone.

Tables 11 and 12 and Figure 2 describe a pilot study I conducted based on my 2005 recordings. I measured 10 tokens (randomly chosen) of short phrases containing words with Low-High-High sequences. I measured the F0 at 10ms intervals for the entire vowel duration for each vowel. From this, I calculated Δ, the difference in mean F0 between adjacent vowels. In examples such as (*shàah*)*tı̨*, 'you (pl) eat,' where the vowels had coalesced, I divided the vowel halfway through its duration.

Table 11. Calculation of Δ

Group 1	Group 2
(L H) H Δ_1 Δ_3 e.g., (*shène*)*tı̨* 'you eat', (*shèts'e*)*zhe* 'we (pl) eat'	L (H H) Δ_2 Δ_4 e.g., (*shèh*)(*tı̨ ha*) 'I will eat', *shè*(*ts'aah*)*zhe* 'we have eaten'

My hypothesis was that, in accordance with tonal feet, there should be a greater jump in F0 across a foot boundary. To test this, I measured different values, i.e. differences in F0 between adjacent syllables, as shown in Table 11. Of several hypotheses I entertained, the strongest was that Δ_1 should be less than Δ_3, which is to say that, in a (Low-High) High sequence, the first High tone is depressed so much that it is phonetically more like a Low tone. Such an extreme effect cannot be explained by interpolation, and therefore provides strong evience for tonal feet. As shown in Table 12, this prediction is borne out, and is statistically significant.

(2) *Results of Δ calculations*

- Δ_1 was 0.7 times the value of Δ_2, though this was not significant (p=0.189).
- Δ_3 was 4.8 times the value of Δ_4, which was significant (p=0.017).
- **Δ_1 was 0.38 times the value of Δ_3, which was significant (p=0.037).**
- Δ_2 was 2.6 times the value of Δ_4, marginally significant (p=0.061).

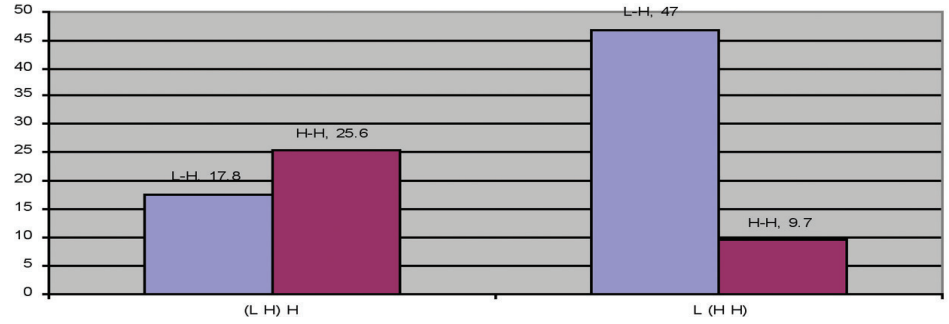


FIGURE 2. Δ-values with different tonal footings

To summarize Table 12 and Figure 2, although the number of tokens was quite small, nevertheless the effect of tonal feet on F0 was categorical and statistically significant. We now turn to the question of how to interpret such data phonologically.[6] Larry Hyman (p.c.) has suggested that in forms such as *shèts'ezhe*, what is happening is actually a *spreading* of the initial Low tone onto the adjacent syllable. This hypothesis is illustrated in Table 13.

TABLE 13. Tone lowering versus tone spreading

	Tone Lowering	Tone Spreading
(L H) H	L M H (*shèts'e*)*zhe*	L H (*shèts'e*)*zhe*
L (H H)	L H H (*shèh*)(*ţi ha*)	L H H (*shèh*)(*ţi ha*)

Under a tone spreading analysis, the reason that the second syllable of *shèneţi* or *shèts'ezhe* shows a depressed F0, compared to other high tones, is that it has in fact been de-linked from its underlying high tone, and associated with the low tone to its left, as shown in Table 13. On the other hand, the reason why these syllables are nevertheless higher than other low tones is that there is a difference in phonetic interpretation between

[6] A reviewer suggests that the phonetic effect shown in (2) and Figure 2 may be due not to lowering H to M due to tonal feet, but rather the *raising* of F0 in stressed syllables. In $(L_1\ H_2)\ H_3$, a raising of the F0 of L_1 would decrease Δ_1, while in $L_1\ (H_2\ H_3)$, a raising of the F0 of H_2 would increase Δ_2. This is indeed a potential confound; more instrumental studies will be necessary to tease out the effects of stress on F0.

two syllables linked to the same low tone, as in Table 13, or two syllables each linked to separate low tones.

In my estimation, this kind of analysis would indeed work to explain tone lowering postlexically, and is even consistent with the existence of tonal feet, insofar as one could say that the tonal foot is the domain of tone spreading in Dogrib. A spreading analysis is not, however, consistent with the types of segmental processes conditioned by tone that we have seen in this paper. Syncope and gemination are due ultimately to sequences of tones which are ill-formed in relation to the metrical structure of the language, but which are not fixed by reassociating tones, but rather by making changes on the segmental tier. If it were possible to reassociate tones and moras in Dogrib, then we would not expect syncope or gemination to happen.

Therefore, I will opt instead for an analysis in which, when the output of the Word Level phonology is (L H) H, the first high tone is lowered to a Mid tone, M. This is illustrated in Table 14.

TABLE 14. H → M postlexically (F0 compression)

/(shèts'e)zhe/	NoContour-Ft	Ident(Tone)
a. L H H | | | (*shets'e*)*zhe*	*!	
b. L L H | | | (*shets'e*)*zhe*		**!
☞c. L M H | | | (*shets'e*)*zhe*		*

In Table 14, I am assuming a gradient scale, whereby a lowing from H to M incurs one violation of Ident(Tone), while H to L lowering incurs two, and is fatal. In other words, tone lowering in Dogrib represents a sort of compromise, whereby postlexically an offending High tone is lowered just enough to make it acceptable with respect to the constraint NoContour-Ft, provided that no other repair strategy was possible at the Word Level or Stem Level.

8. CONCLUSION. In this paper, I have argued that the tonal foot is the driving force behind several seemingly unrelated phonological processes in the Weledeh dialect of Dogrib. Specifically, the constraint NoContour-Ft, which requires that there be no Low-High or High-Low sequences within a foot, is responsible for syncope, gemination, and High to Mid tone lowering. Syncope, in particular, is unusual in this case, since it is conditioned by tone, something which has been claimed to be impossible (Blumenfeld 2006).

It seems, in general, that tones in Dogrib behave like properties of segments, which, from a historical perspective, is not surprising, since until recently they *were* segments, that is, postvocalic glottal elements. Should we, then, believe that syncope, and tonal feet in general, have any synchronic reality at all? Although I believe that syncope could only have arisen under very particular historical circumstances (that is, a weight-conditioned rule that was re-interpreted as tonal), nevertheless this process is still synchronically relevant, since it contributes to a synchronic generalization, NoContour-Ft, which is still productive in the language, since gemination and High to Mid tone lowering still apply productively. If the syncope rule did not apply in Dogrib, there would be, statistically, a much larger number of counterexamples to the generalization that feet should be level, and so learners of the language would have much less evidence that NoContour-Ft is a relevant generalization in the grammar.

In a broader sense, tonal feet in Dogrib illustrate the importance of including low-level allophonic processes in phonological descriptions. There is no necessary connection between being categorical and contrastive. Features which are contrastive in a language can show variation (Anttila 1997), just as features which are non-contrastive can be categorical. In Dogrib, no two lexical items are ever distinguished by consonant length or by High tone versus Mid tone; nevertheless it is essential to look at these phenomena if one is to come up with the right generalizations about Dogrib phonology. Subphonemic details may provide the key to unlocking high-level generalizations about morphophonemics. Tonal feet, therefore, show that it is important to re-examine our assumptions about what types of generalizations qualify as 'phonological', and look at the actual speech signal more closely.

References

Ackroyd, Lynda. 1982. Dogrib grammar. Unpublished ms.

Anttila, Arto. 1997. Deriving variation from grammar. In Franz Hinskens, Roeland Van Hout & W. Leo Wetzels (eds.), *Variation, change, and phonological theory*, 35-68. Amsterdam & Philadelphia: John Benjamins.

Beckman, Jill N. 1995. Shona height harmony: Markedness and positional identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in optimality theory* (University of Massachusetts Occasional Papers in Linguistics 18). 53-76. Amherst: University of Massachusetts.

Bloomfield, Leonard. 1984 (1933). *Language*. Chicago: University of Chicago Press.

Blumenfeld, Lev. 2006. *Constraints on phonological interactions*. Palo Alto, CA: Stanford University dissertation.

Boersma, Paul & David Weenik. 2007. Praat: Doing phonetics by computer. http://www.praat.org (30 January, 2007.)

Borrelli, Doris. 2000. *Raddoppiamento sintattico in Italian: A synchronic and diachronic cross-dialectal study*. Ithaca, NY: Cornell University dissertation.

Casali, Roderic F. 1997. *Resolving hiatus.* New York: Garland Publishing.

Coleman, Phyllis. 1976. *Dogrib phonology*. Iowa City, IA: University of Iowa dissertation.

De Lacy, Paul. 2002. The Interaction of tone and stress in optimality theory. *Phonology* 19. 1-32.

Dresher, B. Elan & Alana Johns. 1995. The law of double consonants in Inuktitut. *Linguistica Atlantica* 17. 79-95.

Flemming, Edward. 1996. *Auditory representations in phonology*. Los Angeles, CA: University of California dissertation.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18. 7-44.

Gil, David. 2001. Escaping Eurocentrism: Fieldwork as a process of unlearning. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork,* 102-132. Cambridge: Cambridge University Press.

Goddard, Pliny Earle. 1912. Analysis of Cold Lake dialect, Chipewyan. *Anthropological Papers of the American Museum of Natural History* 10. 69-170.

Hargus, Sharon. 1988. *The lexical phonology of Sekani*. New York: Garland Publishing.

Hargus, Sharon & Keren Rice. 2005. (eds.), *Athabaskan Prosody*. Amsterdam & Philadelphia: John Benjamins.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Keer, Edward. 1999. *Geminates, OCP, and the nature of Con*. New Brunswick, NJ: Rutgers University dissertation.

Kiparsky, Paul. 1973. Abstractness, opacity, and global rules. In Osamu Fujimura (ed.), *Three dimensions in linguistic theory,* 47-86. Tokyo: TEC.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In Linguistic Society of Korea (ed.), *Linguistics in the morning calm,* 3-91. Seoul: Hanshin.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17(2-4). 351-365.

Krauss, Michael. 2005. Athabaskan Tone. In Sharon Hargus & Keren Rice (eds.). 51-136.
Li, Fang-Kuei. 1933. A list of Chipewyan stems. *International Journal of American Linguistics* 7. 122-151.
Li, Fang-Kuei. 1946. Chipewyan. In Harry Hoijer, Leonard Bloomfield & Mary R. Haas (eds.), *Linguistic structures of Native America* (Viking Fund Publications in Anthropology 6), 398-423. New York: Viking Fund.
Lipscomb, David Robert. 1992. Differences and similarities in the prosodic systems of Western and Eastern Eskimo. *Acta Linguistica Hafniensia* 24. 63-81.
Marinakis, Aliki. 2004. *Seeking simplicity: The preference for minimal syllable structure in Dogrib*. Victoria, BC: University of Victoria MA thesis.
Mohanan, K. P. 1986. *The theory of lexical phonology*. Dordrecht: Kluwer Academic Publishers.
McDonough, Joyce & Peter Ladefoged. 1993. Navajo Stops. *UCLA Working Papers in Phonetics* 84. 151-164.
Myers, Scott. 1997. OCP effects in optimality theory. *Natural Language and Linguistic Theory* 15(4). 847-892.
Pearce, Mary. 2006. The interaction between metrical structure and tone in Kera. *Phonology* 23. 259-286.
Rice, Keren. 1989. *A grammar of Slave*. Berlin: Mouton de Gruyter.
Rice, Keren. 1990. Prosodic constituency in Hare (Athapaskan): Evidence for the foot. *Lingua* 82. 201-244.
Trubetzkoy, Nikolai S. 1969 (1939). *Principles of phonology*. Berkeley: University of California Press.
Tuttle, Siri G. 2005. Duration, intonation, and prominence in Apache. In Sharon Hargus & Keren Rice (eds.). 369-391.
Zec, Draga. 1999. Footed tones and tonal feet: Rhythmic constituency in a pitch-accent language. *Phonology* 16. 225-264.

Alessandro Jaker
amjaker@stanford.edu

10

Language Documentation & Conservation Special Publication No. 2 (May 2010):
Fieldwork and Linguistic Analysis in Indigenous Languages of the Americas,
ed. by Andrea L. Berez, Jean Mulder, and Daisy Rosenblum, pp. 223-261
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4457

Revisiting the source: Dependent verbs in Sierra Popoluca (Mixe-Zoquean)[1]

Lynda Boudreault

The University of Texas at Austin

Sierra Popoluca (SP) is a Mixe-Zoquean language, spoken by about 28,000 individuals in southern Veracruz, Mexico. The objectives of this paper are (1) to explore the structures of dependent verb constructions in SP and the contexts in which they occur and (2) to highlight the stages in which data is gathered and the interplay between text collection, elicitation, and analysis. SP is an ergative, polysynthetic, head-marking language. It has five dependent verb construction types. Early analyses suggested that dependent verbs were non-finite, nominalized forms. Further research indicated that the verbs are components in complex predicates that share inflection for aspect/mood, person, and number. Implicated in the analysis of these constructions are: the prosodic system; the alignment system, which is hierarchically driven with split ergativity; and the number system, also hierarchically driven. The teasing apart of the various grammatical features led to a multi-step process of analyzing and collecting data. By looking at a complex grammatical structure, this paper highlights the interdependency of corpus building, text analysis, and elicitation and the strategies used to negotiate between naturally occurring speech, in which data may be obscured by phonology, and elicited data, which frequently produces periphrastic constructions or alternative utterance types.

1. INTRODUCTION. Sierra Popoluca (SP), also known as Soteapanec (Kaufman 1994, among others), is a Mixe-Zoquean language, spoken in southern Veracruz, Mexico. SP is spoken by 28,194 individuals throughout four municipalities: Soteapan, Tatahuicapan, Hueyapan de Ocampo, and Benito Juarez (INEGI 2000, 2005).[2] It is one of three Gulf Zoquean languages. Its sister languages, Ayapanec and Texistepec, are moribund. Of the three languages, Soteapanec is the only Gulf Zoquean language being learned by children. The analysis presented here comes from data gathered in three communities—Soteapan, Piedra Labrada (Tatahuicapan) and Santa Rosa Cintepec (ten minutes by bus from Los Mangos)—in numerous field visits between 2004 and 2009.

---

[1] My gratitude goes to the speakers of Soteapan, Santa Rosa Cintepec, and Piedra Labrada. The analysis presented in this paper has benefitted greatly from comments from Andrea Berez and Daisy Rosenblum, as well as two anonymous reviewers. It should be noted that all reviewers are not in agreement with certain points in the analysis and all errors are my own. Support for the research presented in this paper comes from grants from Fullbright-Garcia Robles and NSF Doctoral Dissertation Research Grant (BCS-0642425).

[2] 'Sierra Popoluca' is the term commonly used in the communities where Popoluca is spoken. 'Sierra Popoluca' and 'Soteapanec', however, are both exonyms.

The purpose of this paper is two-fold. The first is to explore the structures of dependent verb constructions (DVCs) in SP and the contexts in which they occur. The second is to highlight the stages in which the data have been gathered.

In SP, verbs in the basic clause are obligatorily inflected for aspect or mood (or both) and person. For example in (1),[3] the verb *nɨkk* 'go' is inflected with *ʔa+* '1st absolutive' (subject) and *-pa* 'incompletive' aspect.

(1) *ʔa+nɨktaʔmpa+m* *niʔi=kɨʔim*
ʔa+nɨkk-taʔm-pa+ʔam niʔ=kɨʔim
XB+go-12PL-INC+ALR river=to
'We go to the river.' (MAB.026a)

There are five contexts in which dependent verbs occur; three of which are auxiliary verb constructions (2) and two of which involve temporal/aspectual subordination (3). DVCs are composed of two[4] verbs: the first verb (V1)[5] is inflected for aspect, and the second verb (V2), takes dependent verb morphology and no inflection for aspect or mood. In these sentences the V1 always takes inflection for aspect and mood and may optionally take inflection for person. The V2 is always dependent on the V1 for its aspectual information and obligatorily takes inflection for person.

---

[3] Many of the examples are from texts and are coded to indicate which text (a key to the codes is found in the appendix). When an example is from elicitation, it is coded with the date it was recorded. The orthography used here has the following symbols: y = [j], dy = [dʲ], ty = [tʲ], ch = [tʃ], j = [h], x = [ʃ], tz = [ts], e = [ɛ], o = [ɔ], VV = [V:]. For discussion of phonology I use IPA symbols throughout this paper. The abbreviations used in example glosses this paper are: 1, 2, 3=person; 12PL='1st and 2nd person plural'; 3PL='3rd person plural'; A=Set A person markers; ALR='already'; ANTIP='antipassive'; ASSOC='associative'; AUX='auxiliary'; B=Set B person markers; CAUS='causative'; CMP='completive'; COMP='complementizer'; DEP_{ia}='dependent intransitive-a'; DEP_{ib}='dependent intransitive-b'; DEP_{t}='dependent transitive'; DERIV='derivational morpheme'; DESID='desiderative'; DJO='it is said'; FILL = 'filler'; I='inclusive'; IMP='imperative'; INC='incompletive'; IND='indirective'; INS = 'instrumental applicative'; NEG='negation'; NOM='nominalizer'; OPT='optative'; PASS='passive'; PAST = 'past tense adverb'; PERF='perfect'; PL_{HUM}='plural human'; PL_{NONHUM}='plural nonhuman'; PRO='pronoun'; PROG='progressive'; REP = 'repetitive'; SUBORD='subordinator'; VERS='versive'; X='exclusive'. In glosses, H is an unspecified underlying segment with four surface forms, $-W_2$ is dependent transitive suffix, and $-W_3$ is dependent intransitive type b.

[4] These constructions differ from compounds or serial verb constructions (Crowley 2002; Foley & Olson 1995; Zavala 2000) in that compound verbs in SP are phonologically and morphologically bound and there is no dependent verb morphology. See Boudreault (2009).

[5] I follow Aissen (1994) in labeling the verbs with respect to their position in the construction (i.e. V1 and V2) rather than identifying verbs as 'main', 'auxiliary', or other verbs in a dependency relation to one another.

(2)

	V1	V2		
ʔɨch	*mojo+m*	*ʔa+puʔunyi*	*jeʔm*[6]	*niʔi=kɨʔim*
ʔɨch	moj-W[7]+ʔam	ʔa+puʔn-i	jeʔm	niʔ=kɨʔim
1PRO	begin-CMP+ALR	XB+swim-DEPia	that	water=in

'I began to swim in the water.' (MAB.027)

(3)

ʔi+yóʔy	*jeʔm*	*pakus*	*jeʔm+gamun*
ʔi+yoʔy-W	jeʔm	pak=jos	jeʔm+gak+ʔam+ʔun
3A+jump-CMP	that	trench	that+REP+ALR+DJO

'He jumped the trench there,' (VYT.081)

V1	V2
ʔanhwéj	*mo+ʔi+nyɨ́k*
Ø+ʔanh.wej-W	mo+ʔi+nɨkk-W
3B+shout-CMP	SUBORD+3A+go-DEP_{ib}

'and he ran off mooing (lit. shouting).'[8]

In this paper, I use *DVC* to describe constructions consisting of two verbs in which at least one is syntactically dependent on another for its aspectual information, is inflected with dependent verb morphology, and shares the subject with the verb it depends on.

The verbs in V2 position (the dependent verb) exhibit a number of features that indicate that they pattern like verbs in some contexts and like nouns in others. In fact, early descriptions of Popoluca identified verbs in these contexts as 'gerundial' (Foster & Foster 1948). Work on languages of Mesoamerica has also inquired as to the status of dependent (or non-finite) verbs particularly in Mayan languages, which have been called nominalized verbs or verbal nouns. For example, Larsen (1981:137) states that in Aguacatec "Ss are not cross-referenced by the customary absolutive morphemes but rather by the same ergative prefixes that cross-reference As, thus giving the appearance of nominative/accusative verb agreement system in those environments" and suggests that "derived verbal nouns and, thus the ergative prefixes cross-referencing their respective Ss and As may be viewed as being formally noun possessors." England (2006) has also suggested that non-finite verbs

---

[6] *Je[ʔ]m* 'that' is part of the demonstrative pronoun paradigm *yɨ[ʔ]p* 'this', *je[ʔ]m* 'that', *pe[ʔ]m* 'that yonder'. Nevertheless, usage in texts indicates that *je[ʔ]m* is undergoing reanalysis as a determiner. Because it preserves its function as demonstrative pronoun, I have chosen to use the glossing convention 'that' for the sake of consistency.

[7] The symbol -W represents an underlying phonemic segment of an unknown shape. This segment has four allomorphs in its surface form [-wɨ] ~ [ɔ] ~ [u] ~ Ø. Evidence for the Ø realization is the stress system in Sierra Popoluca. For a detailed description, see Boudreault (2009). Further evidence comes from comparative work with other languages in the Mixe-Zoquean language family (Kaufman 1963, 1997, 2005). The shape of the completive has been reconstructed as *-wɨ in proto-MZ.

[8] *ʔanhwej,* which literally means 'to scream or shout', was translated by the speaker as *bramar* 'to moo (also howl)'. I preserved the speaker's Spanish translation in the English.

in Mam are 'verbal nouns'. Vasquez (2007) has examined non-finite verbs in dependent verb positions in Chol, which do not resemble verbs in terms of focus or modification despite manifesting a number of properties associated with nouns. Mateo-Toledo (2006) has also questioned whether it is possible to analyze non-finite clauses as nominalized forms in his work on Q'anjob'al. In more recent work, Mateo-Toledo (2008) describes verbs in these contexts as 'aspectless' (following in the descriptive Mayanist tradition, citing Craig (1977), England (1983), Zavala (1992), among others) and argues that they form part of complex predicate constructions that have a single value of time, aspect, and modality, and a single set of arguments.

My primary objective here is to show differences between verbs in independent clauses and dependent verbs in SP. The secondary objective is to show the unique properties associated with the different construction types, some of which exhibit features associated with nouns with respect to inflection for person and number. A third objective is to show that while some constructions exhibit some properties associated with nouns, DVCs in each construction type preserve their verbal properties in most contexts. My final goal is to demonstrate the interdependent processes of data collection and analysis by addressing how different observations emerge at different stages of analysis. The final goal essentially recounts the process through which the complete story has been obtained.

Taken as a whole, DVCs pose numerous puzzles with respect to defining categories. In SP, analyzing DVCs has led to regularly revisiting the data, as well as the field, and teasing apart the various interrelated aspects of the grammar to get the complete story. In order to accomplish the objectives of this paper, in section 2 I begin by laying out a brief description of verbs in simple clauses and their associated morphosyntax. The main body of this paper (section 3) lays out the contexts in which dependent verbs occur, the characteristics associated with dependent verbs, and how they differ from verbs in simple clauses. Finally in section 4, I present my methodology and address the evolution that the analysis has undergone, demonstrating how different aspects of the grammar were broken down and investigated over a number of field visits. Presenting the independent stages of analysis is significant, because it illustrates that dependent verbs cannot be taken as a whole, but rather require an understanding of different aspects of SP morphophonemics and morphosyntax.

2. TYPOLOGICAL OVERVIEW AND CHARACTERISTICS OF THE BASIC CLAUSE. SP is an ergative, head-marking language (with split ergative alignment in dependent clauses). It is a polysynthetic and highly agglutinating language. In (4), the subject of the clause *jeʔm pɨ̈ɨxiny* 'the man' is optionally expressed and both the adverb and the O[9] have been incorporated into the verb. SP is predominantly verb-initial, although entire sentences can be communicated with a single word, particularly in naturally occurring speech if a nominal referent is expressed in a preceding sentence (see Payne 1986). When arguments precede the verb, it is pragmatically motivated.

[9] Following Dixon (1979, 1994), I adopt the convention of representing subject, agent, and object as S, A, and O, respectively.

(4) *ʔanh+jeʔega=jukti=ʔak+núʔkpa* *(jeʔm* *pɨɨxiny)*
Ø+ʔanh+jeʔek.ʔaH=jukti=ʔak+nuʔk-pa (jeʔm pɨɨxiny)
3B+quickly=hearth=CAUS+arrive-INC (that man)
'Hurredly, the man got his fire going.' (ESK.063)

2.1. CHARACTERISTICS OF THE VERB IN SIMPLE CLAUSES. In SP, the independent clause consists of a predicate. Verbal predicates minimally take inflection for person and for aspect and/or mood, as shown in (5).

(5) *dya* *dya* *ʔa+monhtoʔoba*
dya dya ʔa+monh-toʔ-pa
NEG NEG XB+sleep-DESID-INC
'No, no I'm not sleeping.' (VYT.016b)

Person is marked with proclitics.[10] Set A clitics mark A, and Set B clitics mark S and O. The person marking proclitics are shown in Table 1.

[10] The term clitic is used to refer to morphemes that are extrametrical (do not take stress) and that occur with different word classes, following Klavans (1985), Zwicky (1977), and Zwicky & Pullum (1983). These characteristics are shown in (i) for enclitics and (ii) for proclitics. The penultimate syllable receives primary stress (unless the final syllable is heavy, e.g. CVːC; CVC is not heavy) and the leftmost syllable of the stressable word (following clitics) receives secondary stress (the phonemic transcriptions are in IPA).

(i) *ʔɨch ʔa+pɨɨxiny+tam* [ʔɨch ʔaˈpiː.ʃɨŋ.tʸaᵑ̩m] 'We are men.' (20070720anonS10)
ʔi+jootʔonh+yaj+nam [ʔiˈjoː.doŋ.jah.nam] 'They still knew.' (GU1.002)

(ii) *ʔanh+ku+jeep-ʔaʔy-pa* [ʔaŋ.kuˌhɛʔːˈbaʔj.paʔ] 'I'm going to scale (the fish).' (20070719anonS5)
na+ku+wih-ʔaʔy-taH-pa [na.kuˌwiʔ.haʔjˈtʸaːpʰ] 'It unravels itself.' (20070603jafs55)

TABLE 1. Person Inflectional Proclitics in Sierra Popoluca[11]

	Set A	Set B
Exclusive	*ʔan+*	*ʔa+*
Inclusive	*tan+*	*ta+*
2nd	*ʔin+*	*mi+*
3rd	*ʔi+*	*Ø+*
1st > 2nd	*man+*	
2nd > 1st	*ʔan+*	

SP exhibits a hierarchical system[12] (Silverstein 1976) in which only higher ranking participants are marked on the verb, and as such only one clitic appears on the verb (Boudreault 2006, 2009). The verb is inflected with Set A markers when A is 1st, 2nd, or 3rd person and O is 3rd person, indicative of a DIRECT alignment pattern. (6) shows a 2nd person A marked with 2nd person Set A *ʔin+*. When A is 3rd person and O is 1st or 2nd person, only Set B clitics for 1st and 2nd person appear on the verb, indicative of INVERSE alignment.[13] In (7), the agent of the derived transitive verb *pak.kaʔ* 'throw off' is the exclusive 1st person, which is marked on the verb with the Set B proclitic *ʔa+*. A third pattern, referred to as the LOCAL configuration by Algonquianists (Hockett 1966), is marked with proclitics that encode speech act participants (SAP)—the speaker (1st person) and the hearer (2nd person)—acting on one another. When A is 2nd person and O is 1st person, the morpheme *ʔan+* is marked on the verb (8). When A is 1st person and O is 2nd person, the transitive verb is marked with the proclitic *man+* (9).

[11] Note that Set A morphemes are used to indicate the agent of a transitive verb, the subject of an intransitive verb, and the possessor of a noun. In an earlier draft of this paper, I distinguish between nouns and verbs by glossing Set A morphemes as ERG when they occur on verbs and PSR when they occur on nouns. This is strictly a glossing convention and does not indicate that person markers and possessive proclitics are homophonous. Here I follow in the Mayanist tradition, labeling Set A for ergative and possessor and Set B for absolutive.

[12] This hierarchical system (Silverstein 1976) is common to Mixe-Zoquean languages. The person marking system reflects a ranking in which speech act participants (1st and 2nd person) rank higher than non-speech act participants (3rd person). This has been noted for the Mixe-Zoque languages Texistepec (Reilly 2002), Chiapas Zoque (Faarlund 2004), Olutec (Zavala 2004), and San Miguel Chimalapa Zoque (Johnson 2000).

[13] An alternative explanation based on the morphophonemics of Sierra Popoluca is possible, although it does not account for the local alignment. In Boudreault (2006) I present the existing analyses—morphosyntactic and morphophonemic—to show that both analyses are not only possible, but they both reinforce the hierarchical system. This analysis is supported by evidence from other languages in the Mixe-Zoque family, which show similar patterns.

(6) 2nd agent > 3rd object (direct):
ʔiny+nyoʔoba ***ʔiny****+wɨdyaaya*
ʔin+noʔ-pa **ʔin**+wɨty=ʔaaya
2A+burn-INC **2A**+husband
'You'll burn your husband.' (Comal.026)

(7) 3rd agent > Exclusive object (inverse):
ʔɨch dya ***ʔa****+pakkáʔ* *maachu*
ʔɨch dya **ʔa**+pak.kaʔ-wɨ maachu
1PRO NEG **XB**+throw.off-CMP macho
'The macho (horse) didn't throw me.' (VVA.017)

(8) 2nd agent > 1st object (local):
tyi+ʔiga ***ʔan****+yaʔachwát?*
tyi+ʔiga **ʔan**+yaʔach=wat-W
why **2>1**+suffer=make-CMP
'Why did you make me suffer?' (ESK.130)

(9) 1st agent > 2nd object (local):
ʔɨch dya ***man****+yaʔachwát*
ʔɨch dya **man**+yaʔach=wat-W
1PRO NEG **1>2**+suffer=make-CMP
'I didn't make you suffer.' (ESK.132)

The distribution of the alignment pattern is shown in Table 2.

Table 2. Alignment Pattern Distribution in Sierra Popoluca

direct	1, 2 > 3
inverse	3 > 1, 2, 3
local	1 > 2, 2 > 1

Verbs in simple clauses are inflected with aspect and/or mood with the suffixes: completive *-W*,[14] incompletive *-pa*, perfect *-neʔ*, imperative *-iʔ*, optative *-ʔiny*, frustrative *-tɨʔp*, and desiderative *-toʔ*. In (10)[15] the verb *ʔak+nuʔk* 'gather, put together' is inflected

[14] See footnote 2 for description of underlying *W* segment.

[15] Tense in Sierra Popoluca is conveyed lexically or indicated through context. Aspect and mood are morphologically expressed. INCOMPLETIVE indicates events that are habitual or not completed; COMPLETIVE indicates events that are completed regardless of the time that the event occurred. It is possible to have an event in the incompletive aspect occur in the past, the present, or the future. In

with the *-W* completive and the verb *nuʔk* 'arrive' is inflected with the incompletive *-pa* and the desiderative *-toʔ*.

(10) incompletive:

jesɨg	*ʔɨch*	*ʔarak+nuʔk**u**+m*	*ʔan+jukti*
jesik	ʔɨch	ʔan+ʔak+nuʔk-**W**+ʔam	ʔan+jukti
then	1PRO	XA+CAUS+arrive-**CMP**+ALR	XA+fire

'Then I put together my fire,'

pero	*dya*	*nuʔk**toʔoba***
pero	dya	Ø+nuʔk-**toʔ-pa**
but	NEG	3B+arrive-**DESID-INC**

'but it didn't want to light (lit. come).' (Comal.008)

Another feature of verbs in independent clauses is that number is indicated through agreement suffixes on the verb: *-taʔm* indicates plurality of the arguments of the verb if the argument is 1st or 2nd ((11) and (12)), and *-yaj* indicates plurality of the arguments of the verb if the argument is 3rd person (13). (Plural marking on the verb does not indicate event plurality.) The plural suffixes are stress bearing inflectional morphemes that precede inflection for aspect or mood.

(11) 1st person plural:

*ʔarak+wiʔk**táʔm**pa*	*ʔan+weewej*
ʔan+ʔak+wiʔk-**taʔm**-pa	ʔan+weewej
XA+CAUS+eat-**12PL**-INC	XA+grandfather
'We fed my grandfather.'	(MAB.038b)

(12) 2nd person plural:

ʔentonsej	*ʔinhku+màtonh**táʔm**um*
ʔentonses	ʔin+ku+matonh-**taʔm**-W+ʔam
then	2A+hear-**12PL**-CMP+ALR

'So, you've all heard'

yɨʔp	*pɨixiny+tyam*	*nɨm**yáj**pa*
yɨʔp	pɨixiny+tam	Ø+nɨm-**yaj**-pa
this	man+PL$_{HUM}$	3B+say- **3PL**-INC

'[what] these men say.' (CP5.002)

---

the anecdote from which the example was taken, the speaker refers to events that occurred in the past and translated these events into Spanish as such. (I later translated the Spanish into English.) Tense in this case was established by context.

(13) 3rd person plural:

peeroj	*ʔagi+tzɨ̀kso ʔps**yáj***	*jeʔm*	*piiyuj*
pero	ʔagi+ʔi+tzɨk=soʔps-**yaj**-W	jeʔm	piiyuj
but	much+3A+CAUS=tire-**3PL**-CMP	that	chicken

'But boy did they tire out that chicken.' (PQH.014)

2.2. CHARACTERISTICS OF THE NOUN. Nouns in SP may occur as arguments of verbs or as the predicate in non-verbal predicate constructions. When they occur as nouns, they may be inflected with Set A clitics to indicate the possessor, as shown in (14). When nouns occur as non-verbal predicates, they are inflected with Set B clitics to mark S, as shown in (15).

(14) *ʔan+pɨ́k* *jeʔm* ***ʔan**+ʔaganh*
ʔan+pɨk-wɨ jeʔm **ʔan**+ʔakʔanh
XA+grab-CMP that **XA**+griddle
'I grabbed my griddle.' (CML.018)

(15) *ʔa* *mich* ***mi**+parteeraj*
ʔa mich **mi**+parteeraj
ah 2PRO **2B**+midwife
'Ah, you're a midwife.' (SoyPartera.001)

Nouns may also be inflected to indicate plurality with the plural enclitics +*tam*, which indicates plurality of human entities (16a), and +*yaj*, which indicates plurality of nonhuman entities (16b).

(16) a. *yoomo+**tam*** 'women' *yoomo* 'woman'
*tzɨ̈xi+**tam*** 'children' *tzɨ̈xi* 'child'

b. *chimpa+**yaj*** 'dogs' *chimpa* 'dog'
tzaʔ+**yaj** 'rocks' *tzaʔ* 'rock'

The plural markers may also agree with the possessor of the noun. In this case, the markers agree with respect to whether the possessor is a SAP or a nonSAP. In (17a), the clitic that appears on the human noun is +*tam*, which agrees with the human noun. In (17b) the noun is inflected with the marker +*yaj,* which in this case agrees with the nonSAP possessor. It cannot agree with the noun, which is human. The examples in (18) illustrate possessed non-human nouns. In (18a) the plural marker agrees with the noun in terms of its non-human property. It may not agree with the 1st person possessor because it is a 3rd person referent. In (18b) the plural marker +*tam* may only agree with the 1st person possessor; it may not agree with the non-human noun. It is also important to note that plural marking is not obligatory, and therefore, in (18) the plural enclitic agrees with the possessor indicating that the possessor is plural; however, the plurality of *chimpa* 'dog' is ambiguous.

(17) a. *ʔi+yom=tɨ́ɨwɨ+**tam***
ʔi+yoomo=tɨɨwɨ+**tam**
3A+woman=sibling+**PL**$_{\text{HUM}}$
'his sisters' / 'their sisters' / *'their sister'

b. *ʔi+yom=tɨ́ɨwɨ+**yaj***
ʔi+yoomo=tɨɨwɨ+**yaj**
3A+woman=sibling+**3PL**
'their sister' / 'their sisters' / *'his sisters'

(18) a. *ʔan+chimpa+**yaj***
XA+dog+**PL**$_{\text{NONHUM}}$
'my dogs' / *'our dog'

b. *ʔan+chimpa+**tam***
XA+dog+**12PL**
'our dog(s)' / *'my dogs'

Finally, when nouns occur as non-verbal predicates, plural markers agree with S. In (19a) the human noun *yoomo* 'woman' is inflected with 1st person and the plural marker *+tam,* which agrees with the 1st person. In (19b) *yoomo* 'woman' is inflected with 3rd person and the plural enclitic *+yaj*, which agrees with the 3rd person subject.

(19) a. *ʔich ʔa+yoomo+tam*
1PRO XB+woman+12PL
'We are women.' (20070704 JAFS13)

b. *jeʔ Ø+yoomo+yaj*
3PRO 3B+woman+3PL
'They are women.'

2.3. SUMMARY OF THE BASIC CLAUSE AND MORPHOLOGICAL CHARACTERISTICS OF VERBS AND NOUNS. Verbs are generally defined as heading the verb phrase and taking inflection for person, aspect, and optionally for number. The alignment system is ergative: Set A marks A and Set B marks S and O. Nouns may head the noun phrase or occur as non-verbal predicates. They take inflection for person: Set A marks possessors and Set B marks S. They are optionally inflected to indicate plurality of the noun or number agreement with the possessor or the S of non-verbal predicates.

3. THE DEPENDENT VERB. In SP, DVCs are those that do not receive inflection for aspect or mood, are dependent on another verb for aspect/mood, and share the subject with the verb that is inflected for aspect or mood. In DVCs, the verb inflected with aspect occurs

in V1 position and the dependent verb occurs in V2 position, with one exception (*mo* subordinator clauses discussed in section 3.1.4). The V1 is always marked with aspect/mood inflection, and the V2 is marked with dependent verb morphology, person marking and plural marking. In (20) the V1 is the auxiliary verb *yaj* 'finish', which is inflected with the incompletive morpheme *-pa*. The V2 is the verb *chinh* 'bathe', which is inflected with the Set B exclusive proclitic *ʔa+*, the dependent suffix *-i* and the plural enclitic *+tam*. In addition, the person may be marked on the V1, and it always agrees with the subject marked on the V2 (21).

(20) *yáj**pa**+m* ***ʔa**+chíinh**i**+**tyam***
yaj-**pa**+ʔam **ʔa**+chiinh-**i+tam**
finish$_{AUX}$-INC+ALR XB+bathe-**DEP$_{ia}$+12PL**
'We finished bathing.' (MAB.031b)

(21) *mich dya+m* ***mi**+ʔóy* ***ʔiny**+ʔáʔm* *ʔiny+choomo*
mich dya+ʔam **mi**+ʔoy-W **ʔin**+ʔaʔm-**W$_2$** ʔin+choomo
2PRO NEG+ALR **2B**+go/return$_{AUX}$-CMP **2A**+see-**DEP$_t$** 2A+grandmother
'You didn't go see your grandmother.' (VVA.040)

Depending on the type of DVC, the dependent verbs in V2 position may be overtly inflected with dependent verb morphology or appear Ø-marked. For example in (22) the verb *mɨ̈ich* 'play' is inflected with the dependent suffix *-i*. (23) shows a construction in which the dependent verb in V2 is transitive and has no overt dependent morphology.

(22) *ʔodoy nɨ́giny mɨ̈ichi*
ʔodoy nɨkk-ʔiny Ø+mɨ̈ich-i
NEG go$_{AUX}$-OPT 3B+play-DEP$_{ia}$
'They shouldn't go play.' (CQS.005)

(23) *ʔestej yɨʔp tuunh+gak sɨʔp ʔi+wɨ̀ɨkuyujkáʔ*
ʔeste yɨʔp tuunh+gak sɨʔ-pa ʔi+wɨ̈=ku+yuj-kaʔ-W$_2$
FILL this other walk$_{AUX}$-INC 3A+good=learn-INS-DEP$_t$
'This other one, he's learning it well.' (CP5.004)

We know there is a segment underlyingly present because stress falls on the final syllable.[16] Establishing dependent morphology in SP is problematic because in a number of the construction types verbs appear bare. Constructions in which the verb in V2 position is intransitive following a small set of auxiliary verbs are inflected with the dependent intransitive suffix *-i*. However, in all constructions in which the verb in V2 position is transitive, the verb is Ø-marked for dependent morphology. As with the completive suffix *-W*, there is evidence to posit an underlying segment. SP prosody and comparative analysis

[16] See fn. 7 above. See fn. 10 for stress rules. See Boudreault (2009) for a description of phonology.

suggest that there is an underlying segment; however the shape of this segment is unknown. Following Kaufman (1997), I assign the label *-W* to account for the presence of said suffix in the morpheme breakdowns.

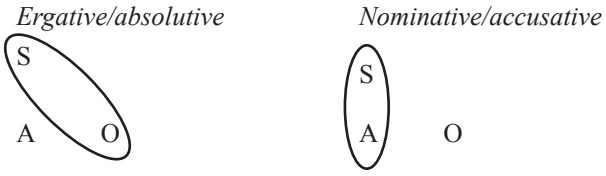
In a number of DVCs, when the dependent verb is intransitive, S is marked with Set A person markers as well as being Ø-marked as a dependent. The pair of examples in (24) and (25) illustrate the intransitive verb *put* 'exit' in independent and dependent contexts, respectively. (24) shows the verb inflected with a Set B person marker in the completive aspect. (25) shows the same verb in V2 position—inflected with a Set A person marker—in a dependent relation to the V1 and inflected for perfect and completive aspect.

(24) *ʔɨch* *ʔa+putu+m.*
ʔɨch ʔa+**put**-W+ʔam
1PRO XB+**exit**-CMP+ALR
'I already went out.' (PQ2.109b)

(25) *dya* *ʔa+joʔynyéʔu+m* *ʔam+**pút***
dya ʔa+joʔy-neʔ–W+ʔam ʔan+**put**-W_3
NEG XB+be.angry-PERF-CMP+ALR XA+**exit**-DEP_{ib}
'I wasn't angry when I left.' (060722ERG061)

This alignment pattern corresponds to a Split-S system conditioned by subordination (Dixon 1994:71, 101-4). That is, independent verbs are marked with Set A person markers to indicate A and Set B person markers to indicate S and O (an ergative/absolutive pattern). Dependent verbs in most contexts, however, are marked with Set A person markers to indicate A and S and Set B to indicate O (a nominative/accusative pattern). The pattern is shown in Figure 1.

FIGURE 1. Ergative/absolutive and nominative/accusative alignment systems.



In DVCs, inflection for aspect/mood and dependent marking are independent of one another.[17] The dependent morphology of the dependent verb is determined by the

[17] Within the Mixe-Zoquean language family, this characteristic is unique to Sierra Popoluca (as noted by Wichmann 1995:103). The person marking in Sierra Popoluca distinguishes between transitive and intransitive, whereas in the other languages of the family dependent marking corresponds with aspect.

type of subordination and the transitivity of the verb. The following examples illustrate constructions with the V1 in completive aspect (26), incompletive aspect (27), and optative mood (28). (26a) shows an intransitive dependent verb marked with *-i*, (26b) shows a transitive dependent verb Ø-marked but bearing person marking, which is shared with the V1, and (26c) shows an intransitive dependent verb Ø-marked as dependent but S is referenced with a Set A person marker. Illustrating incompletive aspect, (27a) shows an intransitive dependent verb inflected with *-i,* (27b) shows a transitive dependent verb Ø-marked as dependent, and (27c) shows an intransitive dependent verb Ø-marked as dependent but whose S is referenced with a Set A person marker. Finally, to illustrate inflection for mood, (28a) shows an intransitive dependent verb with *-i* and (28b) shows a transitive dependent verb that is Ø-marked and bearing person marking referencing the S shared by the V1 and V2. (There are no combinations of optative V1 with dependent clauses showing split ergativity, although other moods are attested.)

(26) completive aspect:

a. *moj**o**+m* *toy-**i*** *ʔi+puʔu*
moj-**W**+ʔam Ø+toy-**i** ʔi+puʔu
begin$_{AUX}$-**CMP+ALR** 3B+ache-**DEP$_{ia}$** 3A+belly
'Her belly began to hurt.' (SA2.009b)

b. *moj**u**+m* *ʔi+ku+wogáʔy* *jeʔm* *ʔi+jaʔyuk*
moj-**W**+ʔam ʔi+ku+wok-ʔaʔy-**W$_2$** jeʔm ʔi+jayʔuk
begin$_{AUX}$-**CMP**+ALR 3A+scold-IND-**DEP$_t$** that 3A+brother
'He began to scold his little brother.' (Anecdota.012)

c. *komo* *dya+m* *wɨʔáaj* *ʔi+yoʔoma=séet*
komo dya+ʔam wɨʔaH-**W** ʔi+yoomo.ʔaH=seet-**W$_3$**
like neg+ALR be.able$_{AUX}$-**CMP** 3A+woman-VERS=return-**DEP$_{ib}$**
'As she could not transform into woman.' (VYT.109)

(27) incompletive aspect:

a. *ʔii* *jemum* *ʔeste* *moj**pa**+m* *wɨ̈ɨ̈=tzɨʔy**i***
ʔii jemɨʔam ʔeste moj-**pa**+ʔam Ø+wɨ̈ɨ̈=tzɨʔy-**i**
and right.there this begin$_{AUX}$-**INC**+ALR 3B+good=remain-**DEP$_{ia}$**
'And there it begins to turn out well.' (CP2.006)

b. *miny**pa**+m* *ʔan+ʔaʔmtáʔm*
miny-**pa**+ʔam ʔan+ʔaʔm-taʔm-**W$_2$**
come-**INC**+ALR 2>X+see-12PL-**DEP$_t$**
'Are you (two) coming to see me?' (Cangrejo.040)

c. *ʔagi+siʔ**p*** *ʔi+mɨichyáj*
ʔagi+siʔ-**pa** ʔi+mɨich-yaj-**W$_3$**
very+walk$_{AUX}$-**INC** 3A+play-3PL-**DEP$_{ib}$**
'They're playing a lot.' (CQS.013b)

(28) optative mood:

a. *ʔokmɨ* *ʔaranh+wejaʔypa* *nɨg**iny*** *wiʔik**i***
ʔok-mɨ ʔan+ʔanh+wej-ʔaʔy-pa nɨkk-**ʔiny** Ø+wiʔk-**i**
afterwards 3A+shout-IND-INC go$_{AUX}$-**OPT** 3B+eat-**DEP$_{ia}$**
'Afterwards we called to him (to ask) if he was going to eat.' (CNC.054b)

b. *ʔagakuʔaʔmyájpa*
ʔagi+ʔa+ku+ʔaʔm-yaj-pa
very+XB+seek.out-3PL-INC
'They look for me a lot,'

ʔiga+ʔɨch *nɨg**iny*** *ʔarak+poʔoyáj*
ʔiga+ʔɨch nɨkk-**ʔiny** ʔan+ʔak+poʔ-yaj-**W$_2$**
COMP+1PRO go$_{AUX}$-**OPT** XA+CAUS+give.birth-3PL-**DEP$_t$**
'that I go help them give birth.' (Partera.029/30)

3.1. CONTEXTS IN WHICH DEPENDENT VERBS OCCUR. The five contexts in which dependent verbs occur include: with Type I auxiliary verbs, with Type II auxiliary verbs, in temporal/aspectual subordinator *mo* 'when' clauses, in temporal/aspectual with no overt subordinator clauses, and progressive auxiliary *siʔ*. In most of these constructions the alignment pattern appears to manifest what Dixon (1994:71, 101-4) calls a 'split-S' of the type described as 'main' versus 'subordinate'. The distribution of construction types and the associated alignment patterns are shown in Table 3. This section will look at each of the five types of multi-verb constructions (and their subtypes), the conditions under which the alignment systems are ergative or nominative, and the unique properties exhibited by each of these constructions.

TABLE 3. Multi-verb constructions with dependent verbs

I	Auxiliary I (Active)	ergative
	(Passive)	nominative
II	Auxiliary II: Modals	nominative
III	Temporal/aspectual subordinator (without *mo*)	nominative
IV	Temporal/aspectual subordinator (with *mo*)	nominative
V	Progressive auxiliary *sɨʔ*	nominative

3.1.1. AUXILIARY VERBS, TYPE I. One context in which dependent verbs occur is Auxiliary Type I verb constructions. Auxiliary verbs in SP are verbs that belong to a small, fixed set and take inflection for aspect/mood.[18] There are seven Type I auxiliary verbs. These are *nɨkk* 'go', *miny* 'come', *ʔoy* 'go and return', *yaj* 'finish' (29), *moj* 'begin' (30), *kus* 'have enough of VERB' (31), *jaʔy* 'be late to VERB'.

(29) ***yáj**pa+m* *wíʔiki*
yaj-pa+ʔam Ø+wiʔk-i
finish$_{AUX}$-INC+ALR 3B+eat-DEP$_{ia}$
'They were already finished eating.' (ESK.073a)

(30) ***moj**pa+m* *ʔi+jetz=tím* *ʔi+way*
moj-pa+ʔam ʔi+jetz=tim-W$_2$ ʔi+way
begin-INC+ALR 3A+brush=stretch.out-DEP$_t$ 3A+hair
'She begins brushing out her hair.' (VYT.010b)

(31) *tan+yoomo* ***kús**u+m* *jóʔyi*
tan+yoomo **kus**-wɨ+ʔam Ø+joʔy-i
IA+woman **be.enough**$_{AUX}$-CMP+ALR 3B+be.angry-DEP$_{ia}$
'Our lady got sufficiently angry. (She'd had enough.)' (ESK.083)

When auxiliary I verbs occur with intransitive verbs, as in (32), the verb is marked with the suffix *-i* and person is marked with Set B person markers. When the dependent verb is transitive, the verb is Ø-marked as dependent. With respect to person marking, the hierarchical system is preserved. (33) and (34) illustrate the DIRECT and INVERSE configurations, respectively. In (33) A is 1st person and O is 3rd person. A is the higher ranking participant and is therefore marked on the verb. In (34) A is 3rd person and O is

[18] Definitions of auxiliary verbs are generally language-specific (Heine 1993). The criteria for defining auxiliary verbs in Sierra Popoluca meet those described by Steele et al. (1981:21).

1st person, the higher ranking participant; therefore, O is marked on the verb with Set B proclitic *ʔa+*.[19]

(32) *ʔich* *mojo+m* ***ʔa****+puʔunyi* *jeʔm* *niʔi=kiʔim*
ʔich moj-W+ʔam **ʔa**+puʔn-i jeʔm niʔ=kiʔim
1pro begin$_{AUX}$-W+ALR **XB**+swim-DEP$_{ia}$ that water=in
'I began to swim in the river.' (MAB.027)

(33) *ʔabeesej* *dya+tyi* *ʔi+kiʔispa* *nikpa* ***ʔan+ʔáʔm***
ʔabeesej dya+tyiH ʔi+kiʔis-pa nikk-pa **ʔan+ʔaʔm-W$_2$**
ʔabeeseh NEG+what 3A+eat-INC go$_{AUX}$-INC **XA+see-DEP$_t$**
'Sometimes he doesn't eat anything. I'm going to see him.' (CNC.055)

(34) *ʔokmi* *ʔóy* ***ʔa****+**ʔáʔm*** *ʔa+ʔich*
ʔok-mi ʔoy-W **ʔa+ʔaʔm-W$_2$** ʔa+ʔich
afterwards go/return$_{AUX}$-CMP **XB+see-DEP$_t$** XB+1PRO
'Afterwards they went to see me.' (Partera.004)

3.1.1.1. PLURAL MARKING OF AUXILIARY I DEPENDENT VERBS. Dependent verbs take inflection for person and number; however, unlike independent verbs, dependent verbs in Auxiliary I constructions are inflected with plural enclitics, rather than stress bearing suffixes. We saw in section 2 that verbs are inflected to indicate number agreement of their arguments with the plural suffixes *-taʔm* '1st and 2nd person plural' and *–yaj* '3rd person plural suffix'.[20] (11) and (13) are repeated here in (35) and (36).

(35) *ʔarak+wiʔk****táʔm****pa* *ʔan+weewej*
ʔan+ʔak+wiʔk-**taʔm**-pa ʔan+weewej
XA+CAUS+eat-**12PL**-INC XA+grandfather
'We fed my grandfather.' (MAB.038b)

[19] This example is one of few in the corpus of naturally occurring text that shows a higher ranking patient. This story was recorded and transcribed in the summer of 2006, however, its significance was not evident until I began investigating non-finite verbs with the goal of determining their status as compared to finite verbs.

[20] The plural marking pattern observed in SP, described as 'plurality split' by Smith-Stark (1974) and Corbett (2000), distinguishes between SAP/nonSAP and HUMAN/nonHUMAN depending on whether markers agree with nouns, arguments, or possessors (Boudreault 2007).

(36) *peeroj* *ʔagi+tzɨksoʔps**yáj*** *jeʔm* *piiyuj*
pero ʔagi+ʔi+tzɨk=soʔps-**yaj**-W jeʔm piiyuj
but much+3A+CAUS=tire-**3PL**-CMP THAT chicken
'But boy did they tire out that chicken.' (PQH.014)

Dependent verbs in this context mark number agreement with enclitics. In (37) the enclitic *+tam* agrees with the 2nd person S, and in (38) *+yaj* agrees with the 3rd person S.

(37) *ʔóy* ***mi**+mɨichi+**tyam***
ʔoy-W **mi**+mɨich-i+**tam**
go/return$_{AUX}$-W **2B**+play-DEP$_{ia}$+**12PL**
'You (all) went to play.' (VVA.041)

(38) *yaju+m* *wiʔiki+**yaj***
yaj-W+ʔam **Ø**+wiʔk-i+**yaj**
finish$_{AUX}$-CMP+ALR **3B**+eat-DEP$_{ia}$+**3PL**
'They finished eating.' (Cangrejo 012)

To summarize, Auxiliary I (active) constructions occur with a closed set of verbs. Intransitive verbs in V2 position are marked with the dependent suffix *-i* and plurality is indicated with plural marking enclitics. Transitive verbs in V2 position are Ø-marked as dependent and plurality is indicated with number agreement suffixes. Person marking is ergative/absolutive.

3.1.1.2. AUXILIARY I CONSTRUCTIONS AND PASSIVE DEPENDENT VERBS . The dependent verb in V2 position may be passive. In SP, passive constructions are those in which the O of a transitive verb—or the primary object (PO)[21] of a ditransitive verb (Dryer 1986)—are advanced to S. The passive suffix is *-taH*. (39) shows the transitive verb *suy* 'to lasso' inflected with the 3rd person Set A clitic to mark A. In (40), the same verb *suy* 'to lasso' is marked with the passive *-taH*. Here S is Ø-marked for 3rd person (Set B).

(39) *ʔi+**xúy*** *jeʔm* *ʔi+wɨyaaya*
ʔi+**suy**-W jeʔm ʔi+wity=ʔaaya
3A+**lasso**-CMP that 3A+husband
'The husband lassoed her.' (VYT082b)

[21] Sierra Popoluca is a primary object (PO) language (Dryer 1986:815). In PO languages, the recipient (benefactive, goal, etc.) of a ditransitive verb is marked on the verb. That is, the recipient shares the object properties of the O of monotransitive verbs; the theme assumes secondary object status. This differs from direct/indirect object languages in that it is the theme that shares the object properties of the O of monotransitive verbs, the recipient is often marked as dative.

(40) ***suy****tyáaj* *jeʔm* *yoomo*
Ø+**suy**-taH-W jeʔm yoomo
3B+**lasso**-PASS-CMP that woman
'The woman was lassoed.' (VYT079)

Auxiliary I constructions are interesting because when the dependent verb is passive, S is marked with Set A person markers. Recall from section 3.1.1 that intransitive dependent verbs mark S with Set B person markers. In (41) the V2 is the transitive verb *toj* 'pay' marked with the passive suffix *-taH*. Yet S in this context is marked with a Set A proclitic. This is the only context in which intransitive dependent verbs in auxiliary I constructions mark S with Set A clitics.

(41) *pues* *nɨkpa* *ʔi+yojtáaj*
pues nɨkk-pa ʔi+yoj-taH-W
then go$_{AUX}$-INC 3A+pay-PASS -DEP$_{ib}$
ʔidyɨk *ʔiga+tzɨɨxi=pinhpa*
ʔityʔɨk ʔiga+Ø+tzɨɨxi=pinh-pa
PAST COMP+3B+child=gather-INC
'She went to be paid to deliver babies (lit. 'to pick babies').' (MAB.274)

In contrast, antipassive verbs in V2 do not exhibit ergative split. In SP, antipassive constructions are those in which the valency of a transitive verb is reduced and only the 'notional' A[22] is expressed on the verb as S. For example, in (42), the transitive verb *wɨɨt* 'massage, rub' is inflected with the Set A clitic *ʔi+*, referencing both an A (the midwife) and O (the woman being massaged). The example in (43) shows the same verb derived with the antipassive suffix *-ʔoʔy* and the verb Ø-marked for 3rd person S. (44) shows the same verb with the antipassive as the dependent of the auxiliary *ʔoy* 'go/return'. Here it is marked with *-i* dependent suffix and Ø-marked for 3rd person S.

(42) *nɨmpa+ʔun* *jesɨk* *ʔi+**wɨɨt**pa*
Ø+nɨm-pa+ʔun jesɨk ʔi+**wɨɨt**-pa
3B+say-INC+DJO when 3A+**massage**-INC
'She says, when she [the midwife] massages [the woman],'

dya+ʔun *ta+nɨmpa* *yɨʔp* *dya* *jeʔ* *tzɨɨxi*
dya+ʔun ta+nɨm-pa yɨʔp dya jeʔ tzɨɨxi
NEG+DJO IB+say-INC this NEG 3PRO child
'it isn't, as we say, this isn't a child.' (GU1.078/9)

[22] Following Gildea (1998).

(43) *ʔii komo tanimpa ʔagi+**wiʔidoʔy**pa ʔi+chíʔ ʔi+xaaja*
ʔii komo ta+nim-pa ʔagi+Ø+**wiɨt-ʔoʔy**-pa ʔi+chiʔ-W ʔi+saaja
and as IB+say-INC much+Ø+**massage-ANTIP**-INC 3A+give-DEP_t 3A+gift
'And because, as we say, she massaged often, they gave her her gift.' (MAB.169)

(44) *ʔóy wiʔid**oʔyi***
ʔoy-W Ø+wiɨt-**ʔoʔy-i**
go_{AUX}-CMP 3B+massage-**ANTIP-DEP_{ia}**
'She went to massage.' (MAB.123)

In this context agreement for number differs from that of inflection of dependent verbs in active Auxiliary I constructions and patterns with number marking of independent verbs. Plural markers in these contexts bear stress, indicating that they are suffixes.

(45) *jaʔypa ʔi+ri+miny**yáj***
jaʔy-pa ʔi+na+miny-**yaj**-W_2
$be.late_{AUX}$-INC 3A+ASSOC+come-**3PL**-DEP_t
'They're late in bringing the chairs.' (20070726rcrS4)

(46) *kus ʔan+ʔuk**táʔm** ʔuunu*
kus-W ʔan+ʔuk-**taʔm**-W_2 ʔuunu
have.enough-CMP XA+drink-**12PL**-DEP_t atole
'We've drunk enough atole.' (20070726rcrS4)

Additional evidence to support the claim that plural inflection in these constructions is with suffixes, rather than enclitics, comes from the order of suffixes. On independent verbs, plural suffixes precede the passive suffix in the verbal template (47). In DVCs, the plural markers also precede the passive suffix (48), demonstrating that plural markers in passive DVCs of Auxiliary I constructions are stress bearing suffixes and not extrametrical clitics.

(47) *ʔa+pajtaʔmtáap*
ʔa+paj-taʔm-taH-pa
XB+enclose-12PL-PASS-INC
'We were locked up.' (CNC.019b)

(48) *ʔantej di kwaatruj diaj ʔo siinkuj*
antes di kwaatruj diaj ʔo siinkuj
before part four days or five
'Four or five days before'

mojpa	*ʔi+k+joodonha**yajtyáa***	*jeentej*
moj-pa	ʔi+ʔak+joodonh-ʔaH-**yaj-taH**-W_2	jeentej
$begin_{AUX}$-INC	3A+CAUS+knowledge-VERS-**3PL-PASS**-DEP_t	people

'he begins to inform the people (the people begin to be informed),'

ʔiga+míny	*mar+ak+joodonhatáʔm*	*ʔɨch*	*ʔam+moʔosba*
ʔiga+miny-W	man+ʔak+joodonh-ʔaH-taʔm-W_2	ʔɨch	ʔan+moʔos-pa
COMP+come-CMP	X>2+CAUS+knowledge-VERS-12PL-DEP_t	1PRO	XA+cook.corn-INC

'"I came to inform you (that) I will cook corn."' (PDLMA.Fiesta.020)

3.1.2. AUXILIARY VERBS, TYPE II. Another context in which dependent verbs occur is with the verbs *wɨʔaaj* /wɨH-ʔaH/ 'to be able to VERB' (49), *jutzaaj* /jutz-ʔaH/ 'be such that VERB' (50), and *ʔanh+jagoʔy* /ʔanh+jak-ʔoʔy/ 'be the first to VERB' (51) .[23] The *wɨʔaaj* and *jutzaaj* constructions have been identified as auxiliary verbs of Type II by Elson (1960a, b), Himes (1997), and Kaufman (1997). In (49), the dependent verb is the intransitive *nay* 'be born'. Notice that the subject, *jeʔm tzɨɨxi* 'the child', is referenced on the verb with the Set A proclitic *ʔi+*. In (50), the dependent verb of *jutz-ʔaH* is the intransitive verb *nuʔk* 'arrive', whose subject is also marked with the exclusive Set A person marker. In (51) *ʔanh+jagoʔy* occurs in V1 position and is inflected with the incompletive. The V2 is the intransitive verb *kaʔ* 'to die', which is marked with a Set A person marker. In Auxiliary II constructions, the V1 may be inflected with person markers (52). Although inflection for person on the V1 is rare in these constructions, when the V1 is inflected, the person markers are Set B.

(49)

***wɨʔaa**+m*	*ʔi+**nyáy***	*jeʔm*	*tzɨɨxi*
wɨʔ-ʔaH-W+tyi+ʔam	ʔi+nay-W3	jeʔm	tzɨɨxi
be.able-CMP+just+ALR	3A+**be.born**-DEP_{ib}	that	child

'It was still possible that the baby will be born.' (PAR.039)

(50)

ʔii	*nɨmyajpa*	*jeʔm*	*pwesteru*	*siʔɨp*	***jutzʔaap***
ʔii	Ø+nɨm-yaj-pa	jeʔm	pwesteru	siʔɨp	**jutz-ʔaH-pa**
and	3B+say-3PL-INC	that	vendors	now	**be.such.that**$_{AUX}$**-INC**

ʔan**+**nuʔktáʔm	*ʔɨch+tyam*	*porkij*	*dya+ʔii*	*ʔanhjakʔoʔoyi*
ʔan+nuʔk-taʔm-W_3	ʔɨch+tam	porkij	dya+ʔiH	ʔanhjak-ʔoʔy-i
XA+arrive-12PL-DEP_{ib}	1PRO+12PL	because	NEG+who	govern-ANTIP-NOM

'And the vendors say, "How can we arrive because there's no president."' (PDLMA.presidente.091)

[23] The verbs *wɨH-ʔaH*, *jutz-ʔaH*, *ʔanh+jagoʔy* are derived verbs that have undergone lexicalization. *wɨH-ʔaH* is derived from the adjective *wɨH* 'good' derived with the versive suffix *-ʔaH*. *Jutz-ʔaH* is derived from *juʔutz* 'how' derived with the versive *-ʔaH*. *ʔanh+jagoʔy* is derived from *juʔutz* 'how'and the versive *-ʔaH*. This requires further research.

(51) *ʔii* *ʔanhjagoʔypa* *ʔi+káʔ*
ʔii ʔanh+jak-ʔoʔy-pa ʔi+kaʔ-W
and be.first-ANTIP-INC 3A+die-DEP_{ib}
'and he'll die first.' (Yerno.016b)

(52) ***ʔa**+wiʔaap* *ʔanh+wíty* *ʔagi+ʔa+tzutpa+m*
ʔa+wiH-ʔaH-pa ʔan+wity-W ʔagi+ʔa+tzut-pa+ʔam
XB+$be.able_{AUX}$-INC XA+walk-DEP_{ib} a.lot+XB+fall-INC+ALR
'I couldn't walk well and I fell a lot.' (PDLMA.rodilla.006)

When the dependent verb is transitive, A is marked with Set A person proclitics and Ø-marked for dependent (53).

(53) *juutza+m* ***ʔiʔ+nyíx***
jutz.ʔaH+ʔam **ʔin+ʔix-W_2**
be.such.that+ALR **2A+see-DEP_t**
'How do you see it?' (SoyPartera.020a)

Like transitive verbs in V2 of Auxiliary I constructions, dependent verbs in modal auxiliary constructions are also observed in the passive. (54) illustrates an Auxiliary Type II construction with a passive V2. Notice in (54) that S is marked with Set A markers.

(54) *dya* *wiʔáap* ***ʔam+metztáaj***
dya wiH-ʔaH-pa **ʔam+metz-taH-W_3**
NEG be.able-INC **XA+look.for-PASS-DEP_{ib}**
'We can't be looked for.' (2070727RCR)

Additionally, the hierarchical system described in section 2 is preserved on the dependent verb. This is illustrated in (55), which shows a higher ranked patient—in this case, the primary object—marked on the verb *seet* 'return' derived with the causative *ʔak+* to mean 'give back'.

(55) *dya* *wiʔaap* ***mi**+k+seedáʔy*
dya wiʔ-ʔaH-pa **mi**+ʔak+seet-ʔaʔy-wi_2
NEG $be.able_{AUX}$-INC **2B**+CAUS+return-IND-DEP_t
'He can't return it to you.' (20070726RCR)

3.1.3. SUBORDINATION WITH NO OVERT SUBORDINATOR. There are two temporal/aspectual contexts in which dependent verbs occur. The first is used to express two events occurring simultaneously. There is no overt subordinator in this construction. The verb in V1 takes inflection for aspect/mood and person (56)-(57). When the V2 is intransitive, the

S of the dependent verb is marked with Set A person markers (56). Note in (57) that both verbs may be transitive and that the V1 takes a Set A person marker.

(56) *dya* *ʔa+joʔynyéʔɨ+m* *ʔam+pút*
dya ʔa+joʔy-neʔ-wɨ+ʔam ʔan+put-W_3
NEG XB+be.angry-PERF-CMP+ALR XA+leave-DEP_{ib}
'I wasn't angry as I left.' (A060722.061)

(57) *yɨkxi* *ʔal.rratuj* *ʔany+nyíx* *minyu+m* *ʔam+miʔɨ*
yɨkxi al.rato ʔan+ʔix-W Ø+miny-W+ʔam ʔan+miʔɨt
like.so later XA+see-CMP 3B+come-CMP+ALR XA+brother-in-law
'A while later I saw my son in law come'

kun *ʔi+yeewa* *ʔi+jɨɨkneʔ* *ʔi+ri+níkk*
con ʔi+yegua ʔi+jɨɨk-neʔ-W ʔi+na+nɨkk-W_2
with 3A+mare 3A+pull-PERF-CMP 3A+ASSOC+go-DEP_t
'with his mare; he was pulling her as he brought her.' (PQ2.056)

These constructions often involve a verb of motion or direction. In most cases the same verbs that occur as an auxiliary in V1 position may occur in V2, however not in the capacity of an auxiliary verb. There is a clear semantic difference in the meaning of the verb when it occurs as an auxiliary in V1 or as a dependent verb in V2. For example, in (58) the auxiliary *miny* 'come' indicates motion towards an event for the purpose of performing the event expressed by the V2. The example in (59) shows the same verb *miny* 'come', however, here the events encoded by the verbs in V1 and V2 position occur simultaneously.

(58) *ʔɨch* ***míny****+am* *ʔa+ʔɨityi+tyam* *yɨʔɨm* *náxwiny*
ʔɨch **miny**-wɨ+ʔam ʔa+ʔity-i+tam yɨʔɨm nax=winy
1PRO **come**-CMP+ALR XB+be-DEP_{ia}+12PL here below
'We came to live down here.' (MAB.174)

(59) *poypa* *ʔi+**míny***
poy-pa ʔi+**miny**-W_2
run-INC 3A-**come**-DEP_{ib}
'It comes running. (It runs as it comes.)' (200707JAF)

3.1.4. SUBORDINATOR MO 'WHEN' CONSTRUCTIONS. The second temporal/aspectual context in which dependent verbs occur is expressed with the subordinator *mo* 'when'. Similar to the subordination constructions described above, the dependent takes person marking and S is marked with Set A clitics. The constructions differ in that the *mo* clause can precede the main verb. *Mo* constructions are used to indicate that an event will occur at the time a second event occurs. (60) shows a *mo* clause in which the dependent *nɨkk* 'go' is inflected with dependent morphology. S in this example is marked on the verb with a Set A person marker. Although any verb may occur as the verb inflected for aspect or the

dependent of the *mo* clause, verbs in these constructions tend to involve a verb of motion or direction.

(60) *ʔi+yóʔy* *jeʔm* *pakus* *jeʔm+ga+m+ʔun*
ʔi+yoʔy-W jeʔm pak=jos jeʔm+gak+ʔam+ʔun
3A+jump.over-CMP that canyon that+also+ALR+DJO

ʔanh+wéj *m+i+nyɨ́k*
ʔanh+wej-W mo+ʔi+nɨkk-W_3
shout-CMP WHEN+3A+go-DEP_{ib}
'He jumped the canyon and shouted when he went.' (VYT.081)

In addition, *mo* clauses may precede or follow the verb inflected for aspect. (61) shows a *mo* clause with the derived verb *mɨj-ʔaH* 'get big' preceding the 'main' clause. The verb of the *mo* clause is marked with dependent morphology—here shown overtly with an allomorph of the shape [wɨ]—and S is marked with a Set A person marker. *Mo* subordinator clauses are the only exception to the rule that the dependent verb occurs in V2 position.

(61) *mo* *ʔi+mɨjaawɨ+m* *nɨmpa*
mo ʔi+mɨj-ʔaH-W_3+ʔam Ø+nɨm-pa
when 3A+big-VERS-DEP_{ib}+ALR 3B+say-INC
'When he grows up, he says'

ʔiga+jeʔ *ʔi+kuʔtpa* *ʔi+jaatunh*
ʔiga+jeʔ ʔi+kuʔt-pa ʔi+jaatunh
COMP+3PRO 3B+eat-INC 3A+father
'that he will eat his father.' (PDLMA.JUU.022)

Finally, passives are also observed in subordinator *mo* constructions. (62) shows the transitive verb *tzam* 'to raise' derived with the passive and inflected with a Set A clitic.

(62) *ʔɨch* *ʔaga+yaʔacháaj* *kwandoj*[24] ***m+an+tzamtáaj***
ʔɨch ʔagi+ʔa+yaʔach-ʔaH-W kwandoj **mo+ʔan+tzam-taH-W_3**
1PRO a.lot+XB+suffer-VERS-CMP when **when+XA+raise-VERS-DEP_{ib}**
'We suffered a lot when we were growing up (lit. being raised).' (7anjatunh.001)

3.1.5. PROGRESSIVE AUXILIARY Sɨʔ. The final context in which dependent verbs occur is with the auxiliary verb *sɨʔ* 'walk'. Like the auxiliary verbs described in section 3.1.1, *sɨʔ* occurs independently as an intransitive verb meaning 'walk' (63). *Sɨʔ* also occurs as an auxiliary verb and indicates progressive aspect. (64) and (65) illustrate *sɨʔ* occurring with transitive and intransitive dependent verbs, respectively.

[24] The use of the Spanish loan word *kwandoj* 'when' serves as a discursive filler in this example. It is rare in these constructions; it is redundant co-occurring with *mo* here.

(63) *jemɨk+piʔk* ***siʔ****iyajpa* *tɨgɨskiʔim*
jemɨk+piʔk Ø+**siʔ**-yaj-pa tɨk=ʔɨskɨ=kɨʔim
there 3B+**walk**-3PL-INC house=behind=at
'There he walks among the houses.' (GU2.105)

(64) *ʔentonses* *mój* *ʔi+m+madáʔy*
ʔentonses moj-W ʔi+ʔanh+mat-ʔaʔy-W
then begin-CMP 3A+speak-IND-CMP
'Then he begins to tell him'

...ʔiga+jeʔm *ʔi+wɨtyyoomo* ***sib*** ***ʔi+péj***
...ʔiga+jeʔm ʔi+wɨty=choomo **siʔ-pa** **ʔi+pej-W**
...COMP+that 3A+wife **PROG$_{AUX}$-INC** **3A+cheat.on-DEP$_t$**
'...that his wife is cheating on him.' (VYT.042a/3)

(65) *dya+m* ***siʔib*** ***ʔi+tyîtz*** *nɨmpa.*
dya+ʔam **siʔ-pa** **ʔi+titz-W** nɨm-pa
NEG+ALR **PROG$_{AUX}$-INC** **3A+dry-DEP$_{ib}$** Ø+say-INC
'"It's not drying," she says.' (SA2.022)

These constructions differ from the auxiliary verb constructions and the temporal subordinator constructions described above in two significant ways. First, unlike the temporal subordinators, *sɨʔ* can occur in V1 or V2 position. Examples are shown in (66) and (67). In (66), the sentence means 'they are playing a lot'. *Sɨʔ* is in V1 position, inflected with the incompletive suffix *-pa*, and the intransitive V2 *mɨɨch* 'play' is inflected with dependent morphology and the Set A proclitic *ʔi+*. In (67), the sentence means 'that's why she's crying'. The verb *wej* 'cry', which is intransitive, is in V1 position, and *sɨʔ* is in V2 position. The V1 is inflected for aspect, the V2 is marked with dependent morphology, and S of the intransitive V2 is marked with 3rd person *ʔi+* (Set A). Therefore, when *sɨʔ* occurs in either position, the V1 takes inflection for aspect/mood, and the V2 marks S with Set A markers.

(66) *ʔagi+**siʔ**p* *ʔi+mɨɨch**yáj***
ʔagi+**siʔ**-pa ʔi+mɨɨch-**yaj**-W$_3$
very+**PROG$_{AUX}$**-INC 3A+play-**3PL**-DEP$_{ib}$
'They are playing a lot.' (CVS.013b)

(67) *jeʔeyukmɨ* *wéjpa* *ʔi+**xîʔ***
jeʔe-yukmɨ wej-pa ʔi+**siʔ**-W$_3$
for.this cry-INC 3A+**PROG$_{AUX}$**- DEP$_{ib}$
'That's why she's crying.' (MAB.019)

The second significant difference between *sɨʔ* constructions and the auxiliary constructions described above is that the meaning conveyed by both orders is the same. In section 3.1.3 we saw that constructions in which the auxiliary verb occurs in V1 position convey motion towards a deictic center to carry out the event expressed by the dependent. (68) shows the auxiliary verb *nɨkk* 'go' in V1 position. When the same verb occurs in V2, the construction conveys that the two events are occurring simultaneously (69).

(68) *ʔan+choomo* ***nɨkpa*** *chiinhi* *nɨʔɨ=kɨʔɨm*
ʔan+choomo **nɨkk**-pa Ø+chinh-i nɨʔ=kɨʔɨm
XA+grandmother **go**$_{AUX}$-INC 3B+bathe-DEP$_{ia}$ river =in
'My grandmother goes to bathe in the river.' (MAB.024)

(69) *ʔɨch* *ʔa+jooppa+m* *ʔan+**nɨk***
ʔɨch ʔa+joop-pa+ʔam ʔan+**nɨkk**-W$_2$
1PRO XB+roll-INC+ALR XA+**go**-DEP$_t$
'I roll as I go along (playing).' (MAB.097)

In *sɨʔ* constructions, however, the meaning conveyed by *sɨʔ* is the same regardless of whether *sɨʔ* occurs in V1 position (70) or in V2 position (71). Notice that in both sentences the V1 takes inflection for aspect, that the V2 takes inflection for person, and that the person marked on the intransitive verb in both cases is Set A.

(70) *sɨʔ**b*** ***any***+*yooxáah*
sɨʔ-**pa** **ʔan**+yooxaH-W$_3$
PROG$_{AUX}$-**INC** **XA**+work-DEP$_{ib}$
'I am working.' (20070706JAF2)

(71) *ʔa+yooxaa**b*** ***an***+*sɨ̂ʔ*
ʔa+yooxaH-**pa** **ʔan**+sɨʔ-W$_3$
XB+work-**INC** **XA**+PROG$_{AUX}$-DEP$_{ib}$
'I am working.' (20070706JAF2)

Sɨʔ constructions manifest similar properties to those described for temporal/aspectual subordination with and without *mo* and auxiliary II constructions with respect to plural marking and passivization. (72) shows a *sɨʔ* construction with plural inflection in which the plural marker *-yaj* is a stress bearing suffix. (73) shows the progressive *sɨʔ* construction in which the transitive verb *pej* 'commit adultery' is passive and S is marked with a Set A marker.

(72) *ʔagi+sɨʔɨb* *i+mɨɨchyáj*
ʔagi+sɨʔ-pa ʔi+mɨɨch-yaj-W$_3$
very+walk$_{AUX}$-INC 3A+play- 3PL-DEP$_{ib}$
'They are playing a lot.' CQS.013b)

(73) *ʔentonses mój* *ʔi+m+madáʔy*
ʔentonses moj-W ʔi+ʔanh+mat-ʔaʔy-W_2
then begin$_{AUX}$-CMP 3A+speak-IND-DEP_t
'Then he began to tell him,'

ʔiga+jesɨk *ta+nimpa* *sɨʔip* ***ʔi+pejtáaj***
ʔiga+jesɨk ta+nɨm-pa sɨʔ-pa **ʔi+pej-taH-W_3**
COMP+then IA+say-INC PROG$_{AUX}$-INC **3A+comit.adultery-PASS-DEP_{ia}**
'that then, as we say, he was being cheated on.' (VYT.096)

3.2. SUMMARY AND DISCUSSION OF DEPENDENT VERBS IN SIERRA POPOLUCA. To summarize, there are five constructions that take dependent verbs. The characteristics shared by each of these constructions are that (1) the V1 takes inflection for aspect/mood, (2) the V2 takes inflection for person and number, and (3) the V2 is marked with dependent morphology. Each of these constructions exhibit unique properties. Auxiliary I active constructions are unique in that the verb in V1 belongs to a closed verb class. The intransitive dependent verbs in V2 are marked with the dependent suffix *-iʔ* and indicate plural agreement with plural marking enclitics. Auxiliary I passive constructions differ from Auxiliary I active constructions in that the subject of a passivized transitive verb is marked with Set A person agreement markers and the dependent morphology is the underlying –W segment. Auxiliary II constructions are unique in that the verb in V1 belongs to a second closed subclass of verb, the dependent in V2 manifests split ergativity in its person marking, and plurality is indicated with plural marking suffixes. Temporal/aspectual *mo* subordinator constructions are unique in that in addition to manifesting split ergativity they employ the subordinator *mo* 'when'. The temporal/aspectual constructions with no overt subordinator are unique in that any verb may occur in V1 and V2 position and no subordinator is required. And finally, *sɨʔ* progressive constructions are unique in that they require the auxiliary verb *sɨʔ*; however, it is not limited to V1 position and may occur in V1 with inflection for aspect or in V2 with inflection for person and number, as well as dependent morphology. I lay out the characteristic properties of each dependent construction in Table 4.

In looking at each of these constructions independently, it is apparent that dependent verbs preserve characteristics of verbs in simple clauses. These characteristics include the passivization of transitive verbs (as well as other derivational processes), plural inflectional morphology, and the hierarchical person marking system. (74) illustrates a transitive root in V2 derived as intransitive with the passive suffix *-taH*.

(74) *yaju+m* *ʔi+kum**táaj*** *jeʔm* *jaam*
yaj-W+m ʔi+kum-**taH**-W_3 jeʔm jaama
finish$_{AUX}$-CMP+ALR 3A+bury-**PASS**-DEP_{ib} that day
'He was finished being burried that day.' (PDLMA.Muerto.035)

(75a and b) also show passive verbs in V2. In (75a) S is plural. Notice that plural inflection precedes the passive suffix, providing further evidence that plural inflectional

morphology is stress bearing (as shown for independent clauses in (47) above). Clitics, however, are extrametrical (non-stress bearing) and occur at the right edge of the word.

(75) a. *ʔantej di kwaatruj diaj ʔo siinkuj*
antes di kwaatruj diaj ʔo siinkuj
before PART four days or five
'Four or five days before'

*mojpa ʔi+k+joodonha**yajtyáa*** *jeentej*
moj-pa ʔi+ʔak+joodonh-ʔaH-**yaj-taH**-W_2 jeentej
$begin_{AUX}$-INC 3A+CAUS+knowledge-VERS-**3PL-PASS**-DEP_t people
'the people begin to be informed (by him),' (PDLMA.Fiesta.020)

b. *ʔiga+míny **mar**+ak+joodonha**táʔm*** *ʔich*
ʔiga+miny-W **man**+ʔak+joodonh-ʔaH-**taʔm**-W_2 ʔich
COMP+come-CMP **x>2**+CAUS+knowledge-VERS-12PL-DEP_t 1PRO

ʔam+moʔosba
ʔan+moʔos-pa
XA+cook.corn-INC
'"I came to inform you all (that) I will cook corn."' (PDLMA.Fiesta.020)

In addition, the person marking hierarchy is preserved in DVCs. (76) shows a dependent verb in which O is the higher ranking participant, and therefore marked on the verb with the Set B marker *mi+*, illustrating an INVERSE configuration. (77) shows a 2nd person A and a 1st person O with the person marker *man+*, illustrating the LOCAL configuration.

(76) *jesɨga ʔukpa mojpa **mi**+maal.mal=nɨʔmáʔy*
jesɨk+ʔiga Ø+ʔuk-pa moj-pa **mi**+mal.mal=nɨm-ʔaʔy-W_2
when+that 3B+drink-INC $begin_{AUX}$-INC **2B**+bad.REDUP=say-IND-DEP_t
'When he's drunk he starts talking bad to you.' (Yerno.004)

(77) *nɨmpa míny **maʔ**+náʔm*
Ø+nɨm-pa miny-W **man**+ʔaʔm-W_3
3B+say-INC $come_{AUX}$-CMP **2>1**+look-DEP_t
'She says, "I've come to see you."' (PAR.005)

Finally, additional evidence indicating that dependent verbs in SP are verbal comes from work on other languages in the Mixe-Zoquean family, which shows that dependent verbs carry aspectual information. It was shown above in (26)–(28) that dependent morphology of dependent verbs is independent of aspect/mood. Auxiliary verbs in San Miguel Chimalapa Zoque also take dependent verbs; however, dependent morphology does distinguish between completive -E and incompletive aspect -wɨ. Dependent marking of depen-

dent verbs in V2 agrees in terms of aspect with the V1: -E if V1 is completive (78);[25] -wɨ if incompletive (79) or in non-declarative mood (80) (i.e. imperative or hortative).

(78) *nək-tam-**wə*** *ʔən+juy-**E*** *boleto*
go-12PL-**COM** 1A+buy-**dCOM** ticket
'We went to buy the tickets.' (Johnson 2000:206)

(79) *jemji* *gaji* *nək-**pa*** *ʔəy-pək=con-**wə***
all there go-**INC** 3A+get=join-**dINC**
'They all go there to receive them.' (Johnson 2000:203)

(80) *min-**ʔo*** *ʔəm+pək=coN-tam-**wə**+ʔam* *haxake+haaʔ*
come-**IMPV2** 2A+get=join-12PL-dINC+NOW female.in.law+NPL2
'Now come meet your mothers-in-law.' (Johnson 2000:209)

The evidence from other languages in the Mixe-Zoquean family shows that while dependent verbs in SP do not convey aspectual information, aspectual marking in V2s does occur in the Mixe-Zoquean family, providing further evidence from a comparative standpoint of the verbal status of independent verbs.

[25] In the San Miguel Chimalapa Zoque examples, the abbreviations are as follows (Johnson 2000): COM, completive suffix; dCOM, dependent completive; dINC, dependent incompletive; IMPV2, imperative; NOW, the equivalent of 'already' in SP; and NPL2, plural marker for pronouns and some nouns.

TABLE 4. Properties of independent and dependent verbs in Sierra Popoluca

Construction Type	Occurs with Aux	Aspect/ mood	Person on V1	Person onV2	Plural suffix	Plural enclitic	Alignment	Subordi-nator	Restrictions
Independent verb	N/A	+	+	N/A	+	-	Erg/Abs	N/A	N/A
V1	N/A	+	depends on type	N/A	-	-	N/A	N/A	N/A
V2									
Aux I, *active*	+	-	optional	+	-	+	Erg/Abs	-	Must be V2
Aux I, *passive*	+	-	optional	+	+	-	Nom/Acc	-	Must be V2
Aux II	+	-	optional	+	+	-	Nom/Acc	-	Must be V2
Subordinator mo 'when'	-	-	oblig. (must agree with S of V1)	+	+	-	Nom/Acc	+	Dependent may precede or follow verb inflected with aspect
Subordination with no overt subordinator	-	-	oblig. (must agree with S of V1)	+	+	-	Nom/Acc	-	Must be V2
si? 'progressive'	+/-	-	optional	+	+	-	Nom/Acc	-	*si?* may be V1 or V2

4. ON THE STATUS OF THE DEPENDENT VERB AND THE EVOLUTION OF THE ANALYSIS. The data on which this analysis is based come from a number of sources. The primary sources consist of transcriptions of narratives and conversations and data from elicitation sessions gathered during a number of field visits ranging from four weeks to 11 months with speakers from the communities of San Pedro Soteapan, Santa Rosa Cintepec, and Piedra Labrada in the state of Veracruz, Mexico. During these field visits, conducted from 2004 through 2009, I have recorded and transcribed narratives and descriptive texts by men and women ranging between the ages of 30 and 70 years. Each field visit consists of conducting text transcription in order to consistently build the corpus of naturally occurring data, as well as elicitation to supplement the analysis of data from texts and complete paradigms in the grammar. The secondary sources consist of materials I obtained from the Project for the Documentation of Languages of Meso-America (PDLMA). These resources include a lexical database (Kaufman & Himes, in preparation), transcribed texts produced by the PDLMA, and texts collected by researchers associated with the Summer Institute of Linguistics. An additional source of textual data on SP is the published text *Hem Tzitzimat* 'La Chichimeca' (Gutiérrez Morales & Wichmann 2001). And finally, examples (text and elicitation) come from Elson (1960a, 1960b), Foster & Foster (1948) and Himes (1997) .

Each of the dependent verb constructions described here has previously been described to some degree, the most detailed and most recent being that of Himes (1997). I have attempted to build on these previous descriptions. The first step in doing so sought to reframe the analysis from the perspective of the dependent verbs and their unifying characteristics; whereas previous analyses characterized constructions based on the auxiliary verbs, lack of auxiliary verbs, and subordinator constructions. The similarity in my approach to my predecessors is that the analysis looks at the language synchronically. Having distinguished the characteristics of each of the constructions independently, a second step and subject of further study in understanding DVCs in SP, as well as those of the languages of the Mixe-Zoquean language family, is to approach the study form a diachronic standpoint.

An important question that arose in describing constructions in which dependent verbs occurred is whether dependent verbs were verbs or nouns, or lay somewhere in between on some continuum. There are two characteristics observed in the early analysis that suggested that dependent verbs were somehow associated with nouns or that verbs in these positions were to some extent nominalized. One observation is associated with plural marking and the other with person marking, specifically that Set A person markers indicate agents of transitive verbs and possessors of nouns. In order to determine the status of dependent verbs, and further investigate plurality and alignment, I needed to focus on each DVC type independently. Upon returning to the field and exploring plurality and alignment with a revised agenda, the observations began to point to dependent verbs preserving verbal characteristics, including derivational properties, the hierarchical system, and plural inflection. Although the position that dependent verbs manifest characteristics of nouns has been revised, it is interesting to note the process through which the analysis went, as well as to demonstrate the interrelated nature of corpus building, text analysis, and elicitation.

4.1. QUESTIONS ABOUT THE STATUS OF DEPENDENT VERBS. The first characteristic suggesting the noun-like behavior of DVCs emerged from an analysis that sought to explain

the Type II auxiliary verbs in terms of complementation (Dixon 2006; Noonan 1985, 2007). The tentative analysis treated *wɨʔ-ʔaH* 'be able to VERB' and *jutz-ʔaH* 'be such that VERB' as complement taking verbs that took subject complements.[26] Set A proclitics, which mark the A of transitive verbs in independent clauses, also mark possessors on nouns. According to the complementation analysis, the examples shown in (81) and (82) (repeated from (49) and (50)) were reanalyzed as: 'The baby's birth was still possible' and 'He knows what our jokes are like'. In this scenario the dependent verb is analyzed as a nominalized verb inflected with possessive person marking and the NP is the subject of the verb.

(81) *wɨʔáa+tyim* *ʔi+náy* *jeʔm* *tzɨɨxi*
wɨʔ-ʔaH-W+tyi+ʔam ʔi+nay-W_2 jeʔm tzɨɨxi
3B+be.able-CMP+JUST+ALR 3A+be.born-DEP_{ib} that child
'The baby's birth was still possible.' (PAR.039)

(82) *jeʔ* *tambyeen* *ʔi+joodonh*
jeʔ tambyeen ʔi+jootonh
3PRO also 3A+know
'He also knows'

jutzáap *tan+moʔogíʔy*
jutz-ʔaH-pa tan+mok-ʔoʔy-W_2
be.such.that-INC IA+joke-DEP_{ib}
'what our jokes are like.' (AVC.016b)

This analysis was taken to the field in the summer of 2007. Evidence indicating that the V2s in these scenarios were not nominalized verbs comes from examples such as the one shown in (83) (repeated from (55)). In this example two characteristics stand out as indicating the verbal status of the V2. First, based on what we know about the hierarchical system, the example shows a higher ranked patient marked on the verb *ʔak+seet* 'return' with a Set B person marker. Second, this example, which was offered spontaneously during elicitation, also demonstrates the extent to which verbs may be derived in the dependent clause. The root verb *seet* 'return, turn around' is intransitive. Here it is derived with the causative *ʔak+*. A secondary object is added with the indirective suffix *-ʔaʔy*. Observations such as these indicate that dependent verbs pattern more like verbs.

(83) *dya* *wɨʔaap* ***mik+seedáʔy***
dya wɨʔ-ʔaH-pa **mi+ʔak**+seet-ʔaʔy-W_2
NEG be.able$_{AUX}$-INC **2B+CAUS**+return-IND-DEP_t
'He can't return it to you.' (20070726 RCR)

[26] *ʔanh+jagoʔy* 'be first to VERB' had not yet been observed at this stage of analysis.

Furthermore, it was noted in section 3.1.2 that person is optionally marked on the V1 in Auxiliary II constructions and that the S of the V1 and V2 are the same. According to the complementation analysis, if the dependent verbs were a possessed nominal S, the V1 *wɨʔaaj* would be Ø-marked for 3rd person. As shown in (84), person marking on V1 and V2 agree.

(84) *ʔa+pɨkpa ʔiga+jeʔ* ***ʔa****+wɨʔaap* ***ʔan****+yooxáaj* *ʔidyɨk*
ʔa+pɨk-pa ʔiga+jeʔ **ʔa**+wɨH.ʔaH-pa **ʔan**+yoox.ʔaH-W_3 ʔityʔɨk
XB+take.wife-INC COMP+3PRO **XB**+be.able$_{AUX}$-INC **XA**+work-DEP$_{ib}$ PAST
'He took me (as wife) because I can work.' (CNC.002b)

The second characteristic suggesting noun-like behavior of dependent verbs comes from plural marking morphology. Nouns may be inflected to indicate plurality of the noun. Plural markers may agree with the noun, with the possessor, or with S when the noun occurs as a non-verbal predicate. As described in section 3.1.1.1, there are two plural enclitics that occur on nouns: *+tam* 'human plural' and *+yaj* 'nonhuman plural'. The fact that enclitics occur on intransitive dependent verbs in the V2 of Auxiliary I (active) constructions suggested that NVPs might pattern like nouns.

(85) *ʔa* *mojo+mun* *ku+pudáʔyi+**yaj***
ʔa moj-W+ʔam+ʔun Ø+ku+put-ʔaʔy-i+**yaj**
ah begin$_{AUX}$-CMP+ALR+DJO 3B+DERIV+exit-IND-DEP$_{ia}$**-3PL**
'They began to escape.' (GU1.123)

Early transcriptions made available provided examples that were inconsistent in marking plurality on dependent verbs, marking some suffixes as clitics and some as stress bearing segments. What was not apparent, although now obvious, was that plural enclitics occurred only on Auxiliary I active verbs, while suffixes occurred everywhere else. In other words, these characteristics did not apply to all dependent verb types. Essentially, what was required was a focused analysis of stress in dependent verbs to determine that there were different patterns.

4.2. THE INTERPLAY BETWEEN TEXT AND ELICITATION. The descriptive work presented here relies first on making observations about what happens in the language, initially on structures that occur in narratives and conversation. Elicitation comes into play in order to verify forms and their uses and to understand what the language does not do, as well as what it does. Understanding the behavior of dependent verbs requires analysis of each of the components that make up these characteristics; some of which appear frequently in texts and some of which require elicitation focusing on paradigms and in some cases attention to the phonology in isolated contexts.

There are several examples that demonstrate the interdependency of corpus building, text analysis, and elicitation, which is crucial to understanding the grammar of the language. One example is illustrated with the *mo* subordinator constructions and the DVCs lacking an overt subordinator, which occur frequently in texts and naturally occurring discourse.

Efforts to elicit these constructions, however, tend to result in periphrastic explanations of the scenarios or the use of the progressive *si?* construction. When these forms do occur in elicitation, they are typically offered spontaneously while working on another aspect of the grammar. Therefore, obtaining examples of the *mo* construction, as well as other subordinator constructions, requires that the text corpus constantly be built and include a range of different genres in which a range of aspectual and temporal information may be expressed.

Similarly, the *jutzaa* 'how is it that, be such that' constructions occur rarely in texts, as well as proving difficult to elicit. *Jutzaa* in texts appears in questions meaning 'how is it that VERB' and declarative statements meaning 'to be such that VERB'. In elicitation, however, only the question form can be obtained. In fact, both meanings tend to be provided with borrowed Spanish expressions in elicitation. In addition, the construction with the Auxiliary II *ʔanh+jagoʔy* 'be the first to VERB' (51) was only observed in a recently transcribed text while writing this paper and has not yet been researched in the field. A large corpus is essential to obtaining a large number of tokens of these constructions.

Elicitation is also invaluable where the data might be obscured by the phonology, especially in cases in which the sentence level prosodic features interacted with word level stress. For instance, it appeared that many dependent verbs are inflected for number agreement with enclitic plural markers, rather than the verbal suffixes that mark independent verbs, which explains inconsistencies observed in a number of text transcriptions. Analysis of these forms was problematic because clitics do not take stress in SP, and they interact with other stress bearing morphemes in verb final position. Isolating DVC construction types in paradigms was necessary to determine that the suffix markers occurring on dependent verbs were in fact suffixes and not clitics. The analysis, however, showed that the plural clitics occur only in intransitive Auxiliary I active dependent verbs and the suffixes occur in the dependent verbs of all other construction types, including Auxiliary I passives. Therefore, understanding the dependent verb template required a clear description of the phonology of independent and dependent verbs.

Research seeking to explicate the structure of the dependent verbs produced more questions, which involved more aspects of the grammar. Returning to the field with new questions and a new lens through which to view the corpus resulted in new observations. The most salient are examples of Auxiliary I constructions in which dependent verbs show ergative alignment. In fact, Auxiliary I constructions were the only multi-verb constructions that showed ergative alignment in the dependent clause, whereas a nominative pattern was observed in all other DVC types.

5. CONCLUSION. My primary objective here has been to show differences between independent and dependent verbs in SP. Verbs in simple clauses are inflected for aspect/mood, number and person, and the alignment system is ergative/absolutive. Although each of these constructions manifests unique properties, the unifying characteristics of the dependent verbs in these constructions are that: (a) they are inflected with dependent verb morphology, (b) they do not take inflection for aspect/mood, (c) they take inflection for person and share the subject with the V1 of the construction, and (d) they take inflection for number.

The second objective is to show the unique properties of each of the DVC types. Dependent verbs occur in five contexts: with Auxiliary I (active and passive), with Auxiliary II, in temporal/aspectual with subordinator *mo* constructions, in temporal/aspectual with no overt subordinator constructions, and with the progressive auxiliary *si?*. The constructions described here differ with respect to the alignment system and plural agreement marking. Like verbs in simple clauses, intransitive dependent verbs of Auxiliary I active constructions show an ergative/absolutive person marking pattern. Unlike independent verbs, intransitive dependent verbs in all other constructions show a nominative/accusative person marking pattern. With respect to plural marking, intransitive dependent verbs of Auxiliary I active constructions mark plurality with clitics, and intransitive dependent verbs of all other construction types mark plurality with stress bearing suffixes.

The third objective is to address observations about the status of dependent verbs, which at early stages of analysis showed characteristics associated with nouns with respect to inflection for person and number, and to demonstrate the interdependent process of data collection and analysis. The discussion presented here demonstrates a predominantly verbal pattern.

Text transcription, data mining, *post hoc* analysis, and controlled elicitation, as demonstrated here, is a cyclic process that looks to data to corroborate predictions driven by linguistic theory and that in turn bear on theory. The analysis presented here illustrates this cycle, providing evidence to show that many independent verb characteristics are preserved on the dependent verb. Some questions that have been raised include: (1) Why do passive dependent verbs in Type I Auxiliary constructions exhibit nominative alignment, while intransitive verbs in Type I Auxiliary (active) constructions do not?; and (2), Why do Auxiliary I (active) constructions differ from all other dependent verb construction in plural inflection? A broader question that emerges is whether these constructions can be reconstructed. Having defined the properties of each of the dependent verbs independently, the next step will consist of comparative analysis with other languages in the Mixe-Zoquean family.

Appendix

The examples in this paper come predominantly from texts I recorded and transcribed with speakers from the communities of Soteapan, Santa Rosa Cintepec, and Piedra Labrada.[27] These texts are archived at the Archive of Indigenous Languages of Latin America at the University of Texas at Austin (http://ailla.utexas.edu).

Code	**Title**	**Speaker**	**Recorded**
Burro	Donkey (personal narrative)	Juliana Albino Franco	02 July 2004
Cangrejo	The crab keeper (local legend)	Juliana Albino Franco	11 July 2007
CNC	When we were married (personal narrative)	Juliana Albino Franco	02 July 2005
Comal	Skillet (personal narrative)	Juliana Albino Franco	19 July 2006
CP5	Construction, part 5 (description)	Braulio Rodrigo Nolasco	10 Sept 2005
CQS	Chaneque (personal narrative)	anonymous	05 May 2005
ESK	Skeleton (local legend)	anonymous	05 May 2005
GU1	Worms, version 1 (local legend)	anonymous	05 May 2005
GU2	Worms, version 2 (local legend)	Eugenia Rodríguez Gutiérrez	11 Nov 2005
MAB	My grandmother (personal narrative)	anonymous	05 May 2005
PAR	Midwife (personal narrative)	Juliana Albino Franco	18 June 2004
viaje	Viaje (PDLMA) (personal narrative)	unknown	unknown
rodilla	Knee (PDLMA) (personal narrative)	unknown	unknown
xuunujti	Xuunujti (SIL) (local legend)	unknown	unknown

[27] Note that where the speaker is listed as anonymous more than one speaker is represented.

PQ2	Broken leg, version 2 (personal narrative)	Juliana Albino Franco	16 July 2006
PQH	Run away chicken (personal narrative)	anonymous	04 April 2005
SA2	Avocado seed, version 2 (ethnomedical desc.)	Juliana Albino Franco	16 July 2006
SoyPartera	'I'm a midwife' (Conversation)	Juliana Albino Franco/ Eugenia Rodríguez Gutiérrez	20 July 2006
VVA	Trip to visit grandma's (personal narrative)	Juliana Albino Franco	02 July 2004
VYT	The cow and the bull (local legend)	anonymous	05 May 2005
Yerno	My son in law (personal narrative)	Juliana Albino Franco	02 July 2005

References

Aissen, Judith. 1994. Tzotzil auxiliaries. *Linguistics* 32. 657-690.

Boudreault, Lynda. 2006. Person marking in Sierra Popoluca: A hierarchical system where phonology matters. Paper presented at the Linguistic Society of America Annual Meeting. Albuquerque.

Boudreault, Lynda. 2007. Plurality splits and hierarchy in Soteapanec (Mixe-Zoquean). Paper presented at the Conference on Indigenous Languages of Latin America III. University of Texas at Austin.

Boudreault, Lynda. 2009. *A grammar of Sierra Popoluca (Soteapanec, Mixe-Zoquean)*. Austin TX: University of Texas dissertation.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Craig, Colette G. 1977. *The structure of Jacaltec*. Austin: University of Texas Press.

Crowley, Terry. 2002. *Serial verbs in Oceanic: A descriptive typology*. Oxford: Oxford University Press.

Dixon, R.M.W. 1979. Ergativity. *Language* 55. 59-138.

Dixon, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Dixon, R.M.W. 2006. *Complementation in typological perspective*. Oxford: Oxford University Press.

Dryer, Matthew S. 1986. Primary objects, secondary objects, and antidative. *Language* 62. 808-845.

Elson, Benjamin F. 1960a. *Gramática Popoluca de la Sierra*. Xalapa: Universidad Veracruzana.

Elson, Benjamin F. 1960b. Sierra Popoluca morphology. *International Journal of American Linguistics* 26. 206-223.

England, Nora C. 1983. Ergativity in Mamean (Mayan) languages. *International Journal of American Linguistic*s 49. 1-19.

England, Nora C. 2006. Cláusulas no finitas en Mam. Paper presented at IX Encuentro Internacional de Linguistica en el Noroeste. Sonora, Mexico

Engel, Ralph, Mary Allhiser de Engel & José Mateo Alvarez. 1987. *Diccionario Zoque de Francisco Leon*. México, D.F.: Instituto Lingüístico de Verano.

Faarlund, Jan Terje. 2004. Person marking in Chiapas Zoque - An optimality account. Paper presented at the Society for the Study of Indigenous Languages of the Americas Annual Meeting. Boston.

Foley, William A. & Mike Olson. 1985. Clausehood and verb serialization. In Johanna Nichols & Anthony C. Woodbury (eds.), *Grammar inside and outside the clause: Some approaches to theory from the field*, 17-67. Cambridge: Cambridge University Press.

Foster, George McClelland & Mary Frazer Le Cron Foster. 1948. *Sierra Popoluca speech* (Smithsonian Institute of Social Anthropology, Vol 8). Washington, DC: Smithsonian.

Gildea, Spike. 1998. *On reconstructing grammar: Comparative Cariban morphosyntax*. Oxford: Oxford University Press.

Gutiérrez Morales, Salome & Søren Wichmann. 2001. Hem ¢i¢imat, 'La Chichimeca'. *Tlalocan* 13. 315-334.

Harrison, Roy, Margaret Harrison & H. Cástulo García. 1981. *Diccionario Zoque de Copainlá*. Mexico, D.F.: Instituto Lingüístico de Verano.

Heine, Bernd. 1993. *Auxiliaries: Cognitive forces and grammaticalization*. New York: Oxford University Press.

Himes, Valerie. 1997. *Multi-verb constructions in Sierra Popoluca*. Pittsburgh, PA: University of Pittsburgh MA thesis.

Hockett, Charles F. 1966. What Algonquian is really like. *International Journal of American Linguistics* 32. 59-73.

Instituto Nacional de Geografía, Estadística e Informática. 2000. *Manual y catálogo para la codificación de las lenguas indígenas, XII Censo General de Población y Vivienda*. México, D.F.

Instituto Nacional de Geografía, Estadística e Informática. 2005. *Manual y catálogo para la codificación de las lenguas indígenas, XII Censo General de Población y Vivienda.* México, D.F.

Johnson, Heidi. 2000. *A grammar of San Miguel Chimalapa Zoque*. Austin, TX: University of Texas dissertation.

Kaufman, Terrence. 1963. Mixe-Zoque diachronic studies. Unpublished ms.

Kaufman, Terrence. 1994. The native language of Meso-America. In Christopher Moseley & R.E. Asher (eds.), *Atlas of the world's languages*, 46-76. London & New York: Routledge.

Kaufman, Terrence. 1997. Grammaticalization through incorporation in the diachrony of Mije-Sokean verbs. Paper presented at The Seventh Biennial Symposium of the Department of Linguistics of Rice University: "On the Interface Between Comparative Linguistics and Grammaticalization Theory: Languages of the Americas". Rice University. Houston.

Kaufman, Terrence. 2005. A typological sketch of the Mije-Sokean languages. Unpublished ms.

Kaufman, Terrence & Valerie Himes. Forthcoming. Soteapanec lexical database. The project for the documentation of the languages of Meso-America. Pittsburgh: University of Pittsburgh.

Klavans, Judith. 1985. The independence of syntax and phonology in cliticization. *Language* 61. 95-120.

Larsen, T.W. 1981. Functional correlates of ergativity in Aguacatec. Paper presented at the Berkeley Linguistic Society. University of California, Berkeley.

Mateo-Toledo, B'alam. 2006. Explorations in nonfinite clauses in Q'anjob'al. Paper presented at the IX Encuentro Internacional de Linguistica en el Noroeste. Sonora, Mexico.

Mateo-Toledo, Eladio. 2008. *The family of complex predicates in Q'anjob'al (Maya): Their syntax and meaning*. Austin, TX: University of Texas dissertation.

Noonan, Michael. 1985. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description*, 42-140. Cambridge: Cambridge University Press.

Noonan, Michael. 2007. Complementation. In Timothy Shopen (ed.), *Language typology and syntactic description,* 2nd edn., 52-150. Cambridge: Cambridge University Press.

Payne, Doris L. 1986. Basic constituent order in Yagua clauses: Implications for word order universals. In Desmond C. Derbyshire & Geoffry K. Pullum (eds.), *Handbook of Amazonian languages*, 407-439. Berlin: Mouton de Gruyter.

Reilly, Ehren. 2002. A survey of Texistepec Popoluca verbal morphology. Unpublished ms.

Silverstein, Michael. 1976. A hierarchy of features and ergativity. In R.M.W. Dixon (ed.), *Grammatical categories in Australian Languages*, 112-171. Canberra: Australian Institute for Aboriginal Studies.

Smith-Stark, Thomas C. 1974. The plurality split. Paper presented at the Chicago Linguistics Society. University of Chicago.

Steele, Susan, Adrian Akmajian, Richard Demers, Eloise Jelinek, Chisato Kitagawa, Richard Oehrle & Thomas Wasow. 1981. *An encyclopedia of AUX: A study in cross-linguistic equivalence*. Cambridge, MA: MIT Press.

Vasquez Alvarez, Juan J. 2007. *Non-finite clauses in Chol*. Austin, TX: University of Texas MA thesis.

Wichmann, Søren. 1995. *The relationship among the Mixe-Zoquean languages of Mexico: Studies in the indigenous languages of the Americas*. Salt Lake City: University of Utah Press.

Zavala, Roberto. 1992. *El Kanjobal de San Miguel Acatán*. México: Universidad Nacional Autónoma de México.

Zavala, Roberto. 2000. *Inversion and other topics in the grammar of Olutec (Mixean)*. Eugene, OR: University of Oregon dissertation.

Zavala, Roberto. 2004. Obviación en oluteco (mixeana). Paper presented at the Conference on Indigenous Languages of Latin-America-I. The University of Texas at Austin.

Zwicky, Arnold M. 1977. On Clitics. Bloomington: Indiana University Linguistics Club.

Zwicky, Arnold M. & Pullum, Geoffrey K. 1983. Cliticization vs. inflection: English *n't*. *Language* 59. 502-513.

Lynda Boudreault
ldejong@mail.utexas.edu

ISBN 978-0-8248-3530-9